

DOCUMENT RESUME

ED 105 369

95

CG 400 121

AUTHOR Galant, Richard, Comp.; Moncrieff, Nancy J., Comp.
TITLE Testing and Testing Programs. Searchlight: Relevant Resources in High Interest Areas. Update Search 14U-2.
INSTITUTION ERIC Clearinghouse on Counseling and Personnel Services, Ann Arbor, Mich.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
BUREAU NO BR-6-2487
PUB DATE Dec 74
CONTRACT OEC-3-6-002487-1579
NOTE 31p.; For related document, see ED 082 45
AVAILABLE FROM Impact Publications, School of Education, Room 2108, The University of Michigan, Ann Arbor, Michigan 48104 (HC \$1.50, cash or money order must accompany request)

EDRS PRICE MF-\$0.76 HC-\$1.95 PLUS POSTAGE
DESCRIPTORS *Annotated Bibliographies; Colleges; *Counseling; Elementary Schools; High Schools; *Literature Reviews; Measurement Instruments; Research Reviews (Publications); *Resource Materials; *Testing Programs; Tests
IDENTIFIERS *Searchlight

ABSTRACT

This updated search of the ERIC system, "Dissertation Abstracts," and the journal literature yielded 153 documents which provide a review of comprehensive testing programs (elementary, senior high school, and college) with special emphasis on specific tests that the practicing counselor may want to utilize. (CJ)

searchlight

Relevant Resources in High Interest Areas

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

14U-2 UPDATE SEARCH

Compiled by Richard Galant
and Nancy J. Moncrieff

December 1974

A review of comprehensive testing programs (elementary, senior high school and college) with special emphasis on specific tests that the practicing counselor may want to utilize.

(153 document abstracts retrieved)

\$1.50

Tests and Testing Programs

ED105369

CG 400 121

Introduction

This information packet, prepared by the ERIC Counseling and Personnel Services Center, is intended to alert the user to a body of literature on a topic of current interest to counselors. It identifies research reports that have been cited in the Educational Resources Information Center's (ERIC) publication, Research in Education (RIE), in Dissertation Abstracts International, and in ERIC's Current Index to Journals in Education (CIJE) from March 1973 through September 1974.

Ordering Instructions

Searchlight has attempted to give availability for all materials listed in this packet. In most cases, it is possible to obtain a personal copy of the title listed. The sources fall into three groupings:

ERIC Documents

References in this search for which an ED (ERIC Document) number is given may be ordered from the ERIC Document Reproduction Service (EDRS). Copies are available in either hard (photo) copy or in microfiche form. The microfiche require a special machine for use. To order any of the ED materials, please refer to the EDRS Order Blank at the back of this packet.

Doctoral Dissertations

All dissertations listed in this search have been drawn from Dissertation Abstracts International, a publication of University Microfilms. They are available on microfilm (MF) at \$5.00 per dissertation, or in bound photo copy (X) at \$11.00 per dissertation from University Microfilms.

To order, give the following information:

- 1) Order number
- 2) Author's last name and initials
- 3) Type of copy desired (microfilm or hard copy)
- 4) Your name and address (list separately for billing and shipping, if they are different)

Send your order, with payment, to University Microfilms, 300 North Zeeb Road, Ann Arbor, Michigan 48106.

Journal Articles

Journal articles are available from the original journal in library or personal collections. Refer to the entry for volume and page designations.

ERIC Documents

ED 069 051

EC 050 086

Cook, John J.
Test Anthology: Fugitive and Standardized Tests.
Wisconsin State Dept. of Public Instruction,
Madison.

Pub Date [71]
Note—257p.

EDRS Price MF-\$0.65 HC-\$9.37

Descriptors—*Annotated Bibliographies,
Behavior Rating Scales, *Exceptional Child
Education, *Handicapped Children, *Stan-
dardized Tests, Student Evaluation, *Testing

The test anthology designed for special educa-
tion administrators, teachers, and program plan-
ners, contains approximately 300 entries consist-
ing of standardized and fugitive tests, behavior
rating scales, and summary evaluation sheets. En-
tries are indexed by key words which are descrip-
tive of the testing instrument and by author. An
alphabetized contents table precedes both key
word and author index. Entries generally in-
dicate key word, title, purpose of the instrument,
author, age range for which the test is designated,
number assigned for computer alphabetizing,
library call number, category or type of instru-
ment, date the instrument was entered into the
index, and all key words for the instrument. It is
explained that special education personnel may
request by mail from the Department of Public
Instruction Library in Madison, Wisconsin, data
sheets which supply ordering information and
pertinent data abstracted from the instrument
manual or instructions. Relevant research and
evaluative studies are included. Data sheets are
intended to allow the user to make a fairly
detailed appraisal of the entry without examining
the instrument. It is noted that copies of the test-
ing instruments may be procured only by contact-
ing the developers directly. (GW)

ED 069 681

TM 002 132

Measures of Self Concept: Grades K-12.
Instructional Objectives Exchange, Los Angeles,
Calif.

Spons Agency—California Univ., Los Angeles.
Center for the Study of Evaluation.
Note—100p.

Available from—Instructional Objectives
Exchange, P.O. box 24095, Los Angeles, Calif.
90024

Document Not Available from EDRS.

Descriptors—*Affective Objectives, *Attitude
Tests, Educational Objectives, Intermediate
Grades, *Measurement Instruments, Primary
Grades, Questionnaires, Secondary Grades,
*Self Concept Tests, Student Attitudes, Tests

Identifiers—Choose a Job Inventory, Class Play,
Parental Approval Index, Self Appraisal Inven-
tory, Television Actors

This collection of instructional objectives and
related measuring instruments is the first of In-
structional Objectives Exchange collections
devoted to affective objectives. It concerns mea-
sures of self concepts. The collection is divided
into five major sections. The discussion of the
collection's organization is followed by a descrip-
tion of the development procedures. The third
sections details possible uses. The objectives
themselves are listed according to grade ranges,
that is primary (grades K-3), intermediate (grades
4-6), and secondary (grades 7-12). At the close
of each objective a page reference refers to the
concluding section which contains measuring
devices, their rationales, and directions for ad-
ministering and scoring them. Measures included
are Self Appraisal Inventories at the primary, in-
termediate and secondary levels; Choose a Job
Inventory; Parental Approval Index; The Class
Play; Television Actors; What Would You Do? at
both intermediate and secondary levels; Work
Posting and Perceived Approval Situations. (DJ)

ED 069 694

TM 002 145

Pedri, Bonnie Pedri, D. T.
Intelligent Intelligence Testing.
Note—7p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Cognitive Processes, Culture Free
Tests, Evaluation Criteria, Intelligence Dif-
ferences, *Intelligence Quotient, *Intelligence
Tests, Measurement Instruments, Measurement
Techniques, *Racial Discrimination, Technical
Reports, *Testing, Test Interpretation
Identifiers—Rosenhal Effect

Intelligence tests should be used to help per-
sons; they should not be used to penalize persons.
Furthermore, our focus should be on treatment; it
should not be on labeling. IQ testers often stig-
matize young children and poor persons (chil-
dren, adolescents, adults). Large groups of Black
Americans, Spanish Americans, and Indian Amer-
icans are probably mis-classified as to ability
because of a differential society and culture. This
paper attempts to develop a better understanding
of tests and testing. If intelligence testing is to
continue, it should be done intelligently. But the
time, effort, and money could be better spent in
treatment procedures: developmental, corrective,
remedial, educational, vocational, personal, so-
cial. (Author)

ED 069 697

TM 002 148

Horne, Eleanor V., Ed.
Test Collection Bulletin.
Educational Testing Service, Princeton, N.J.

Pub Date Jul 72
Note—28p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Academic Achievement, Achieve-
ment Tests, Attitude Tests, *Bibliographies,
Bulletins, *Catalogs, *Measurement Instru-
ments, National Competency Tests, Perception
Tests, Psychological Tests, Questionnaires, Rat-
ing Scales, *Resource Materials, *Tests, Test
Selection

Identifiers—Project Head Start

This is a quarterly digest of information on
tests located in the test collection of Educational
Testing Service. In addition to the extensive lib-
rary of tests and other measurement devices, it also
includes a special Head Start test collection
established to provide information about instru-
ments for those engaged in research or project
direction involving young children. New Acquisi-
tions include sections on achievement; common
examinations; specialty examinations; personality,
interest, attitudes and opinions; and miscellane-
ous, sensory motor, unidentified. A second sec-
tion reports announcements received related to
tests. A third section reports testing programs for
1972-1973, reporting the test, administration, and
contact persons for each. The final section gives
addresses of publishers and organizations whose
materials and services are mentioned in the Bul-
letin. (DJ)

ED 069 722

TM 002 174

Test Analysis Manual.
Escambia County School Board, Pensacola, Fla.
Evaluation Services
Note—23p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Computers, Evaluation, Evaluation
Methods, Information Processing, Instructional
Programs, *Item Analysis, *Manuals, Perfor-
mance Tests, *Services, Standardized Tests,
Student Evaluation, *Testing, Tests

This manual has four purposes. They are (1)
to list test analysis services available to teachers

(2) to explain information on the analysis prin-
touts, (3) to provide help in interpreting analysis
results, and (4) to suggest possible uses of test
analysis data. It is noted that test analyses ser-
vices are available to teachers for teacher-made
tests and surveys if standard answer sheets are
used, as well as for some standardized tests. Any
one or all of the following services may be
selected: (1) class performance analysis, (2) item
analysis, (3) frequency distribution, and/or (4)
response analysis. It is felt that by using the data
generated, teachers will be helped in planning in-
struction, in evaluating student progress, and in
improving test items. (Author)

ED 069 724

TM 002 269

Whitely, Susan S., Davis, Rene V.
A Model for Psychometrically Distinguishing Ap-
titude from Ability.

Minnesota Univ., Minneapolis, Dept. of Psychol-
ogy

Spons Agency—Office of Naval Research,
Washington, D.C. Psychological Sciences Div.

Report No.—MU-TR-3007

Pub Date 25 Jul 72

Note—35p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Ability Identification, *Aptitude
Tests, College Students, Educational Research,
*Mathematical Models, *Measurement
Techniques, *Motor Reactions, *Predictive
Measurement, *Psychometrics, Research
Methodology, Statistical Analysis, Tables
(Data), Technical Reports, Test Results, Test
Validity

It is now widely agreed that current ability
measures reflect a complex interaction of en-
vironment with genetic potential. This leads to a
basic measurement problem since persons with
the same measured ability may vary widely in
potential due to non-equivalent learning opportu-
nities. The purpose of this paper is to present a
model which may hold some promise in
psychometrically distinguishing ability (current
status) from aptitude (potential). Data on a sim-
ple ability are analyzed according to the model to
illustrate how some of the practical problems may
be solved. (Author)

ED 069 795

TM 002 281

Diamond, Esther E.
The Masculinity-Femininity Scale in Interest Mea-
surement: An Idea Whose Time Has Passed.

Pub Date Sep 72

Note—11p.; Paper presented at convention of the
American Psychological Association, (Honolu-
lu, Hawaii, September 1972)

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Females, Males, *Measurement In-
struments, *Occupational Tests, *Psychological
Tests, Role Perception, *Sex Discrimination,
Speeches, *Test Bias

Identifiers—Kuder Occupational Interest Survey,
Kuder Preference Record, Minnesota Voca-
tional Interest Inventory, Strong Vocational In-
terest Blank

The lag that exists between traditional mea-
sures of masculinity and femininity in occupa-
tional interests and the changing role of women
in the world of work is discussed. It is stated that
most masculinity-femininity scales in use today
measure the degree of conformity with socially
and culturally determined sex roles. Scales
discussed are the Strong Vocational Interest
Blank (SVIB), the Kuder Occupational Interest
Survey (OIS), the Minnesota Vocational Interest
Inventory, and the Kuder Preference Record—Or-

occupational norms. What is needed in the way of practice and further research is given as follows: (1) further research on the question of whether separate norms should be developed for the same occupation, on the basis of sex; (2) newer criterion group data should be developed for all inventories; (3) the term Masculinity/Femininity as applied to psychological scales such as measures of interest should be rejected as an idea whose time has definitely passed; and (4) lack of available data should not be used to limit women's or men's career options. (DB)

ED 070 794 UD 013 104

Green, Donald Ross. Draper, John F.
Exploratory Studies of Bias in Achievement Tests.
CTB/McGraw-Hill, Monterey, Calif.
Pub Date Sep 72

Note—59p. Paper presented at the American Psychological Association Annual Convention, Honolulu, Hawaii, September 1972

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Achievement Tests, Bias, Comparative Testing, Educational Testing, Ethnic Groups, *Group Tests, Item Analysis, Minority Groups, *Racial Differences, Research Methodology, *Social Differences, *Test Bias, Testing Problems, Test Validity

Identifiers—California Achievement Tests
This paper considers the question of bias in group administered academic achievement tests, bias which is inherent in the instruments themselves. A body of data on the test of performance of three disadvantaged minority groups—northern, urban black, southern, rural black, and southwestern, Mexican-Americans—as tryout samples in contrast to white, advantaged groups in the same regions, was analyzed using five different general methods for examining tests for bias. In an item tryout, a set of items is administered to a sample of the relevant population and the results are then examined item by item in an effort to pick the more effective items. The first method is an item selection routine using the point biserial correlation for each item as the criterion. The second method, group by score interactions, involves dividing the tryout group into, say, fourths, based on quartiles, and examining the proportion of the cases making each possible response in each of these levels. The third method involves plotting item difficulties so as to locate aberrant items. The fourth method involves estimating and plotting item characteristic curves separately for each group and comparing the plots. The fifth method comprises various intergroup factor analytic approaches. (Author/JM)

ED 070 799 UD 013 110

Williams, Robert L.
The BITCH-100: A Culture-Specific Test.
Washington Univ., St. Louis, Mo.
Spons Agency—National Inst. of Mental Health (DHEW), Bethesda, Md.
Pub Date Sep 72

Note—21p. Paper presented at the American Psychological Association Annual Convention, Honolulu, Hawaii, September 1972

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Achievement Tests, Culture Free Tests, Dropouts, High School Students, Inner City, *Intelligence Tests, Negro Culture, Negro Dialects, Racial Differences, *Test Bias, *Test Construction, Testing Problems, *Test Validity

Identifiers—BITCH 100, Black Intelligence Scale of Cultural Homogeneity, California Achievement Test
The aim of this paper is to describe the rationale and evaluation of the Black Intelligence Scale of Cultural Homogeneity (BITCH). A "culture specific" test is used to determine the taker's ability to function symbolically or to think in terms of his own culture and environment. A combination of dialect specific and culture specific tests would certainly enhance the reliability of measuring what is inside the black head; this is the basic rationale for the BITCH 100.

Over two years, a 100-item test was developed. The purpose of the first experiment was to demonstrate that the test would discriminate black from white takers. One hundred white and 100 black high school students ranging in age from 16 to 18 years, half from low

socioeconomic levels and half from middle income levels, from the city of St. Louis took the BITCH 100. The black group showed a clear superiority over the white group. The distribution of scores approximated a normal distribution in which blacks comprise the upper half, whites the lower half. Twenty-eight black Neighborhood Youth Corps high school "drop outs" were administered the BITCH and the California Achievement Test in the second experiment. The results confirm the hypothesis regarding the sensitivity of the BITCH in picking up "intellectual indicators" not commonly found in conventional tests. (Author/JM)

ED 071 935 SO 005 046

Popham, W. James
An Evaluation Guidebook. A Set of Practical Guidelines for the Educational Evaluator.
Instructional Objectives Exchange, Los Angeles, Calif.

Spons Agency—Bureau of Indian Affairs (Dept. of Interior), Washington, D.C.

Pub Date 71

Note—129p

Available from—Instructional Objectives Exchange, Box 24095, Los Angeles, California 90024 (\$2.50)

Document Not Available from EDRS.

Descriptors—*Behavioral Objectives, Educational Accountability, Educational Improvement, *Educational Objectives, Evaluation Criteria, *Evaluation Techniques, Measurement Goals, *Measurement Techniques

This booklet is designed to assist those responsible for conducting educational evaluations. The work contains a collection of recently developed technical procedures relevant to educational evaluation and to describe them in a straightforward manner. The guidelines are organized into three sections: 1) Objectives, 2) Measurement, and 3) Data Collection and Analysis. In the objectives section, seven guidelines are presented which deal with such matters as how objectives should be specified so that they will be of most utility to the evaluator. The measurement section contains five guidelines treating such problems as how the evaluator should devise measures which satisfactorily assess whether an objective has been achieved. The section dealing with data collection and analysis includes eight guidelines dealing with the types of designs evaluators should employ for various situations and how to treat the data secured via these designs. A summary of all guidelines is presented at the close of these three sections. A final section, Playing by the Rules, provides a fictitious example of an educational evaluator who conducts his/her work consonant with the guidelines. A set of selected references, organized around the major sections of the guidebook concludes the volume. (FDI)

ED 072 096 TM 002 336

Horne, Eleanor V. Ed.
Test Collection Bulletin, Vol. 6, No. 4, October 1972.

Educational Testing Service, Princeton, N.J.

Pub Date Oct 72

Note—27p. Address questions about Head Start to Head Start Test Collection, Educational Testing Service, Princeton, N.J.

Available from—Eleanor V. Horne, Test Collection, Educational Testing Service, Princeton, New Jersey 08540 (subscription \$2.00, foreign readers \$2.50)

EDRS Price MF-\$0.65 HC Not Available from EDRS.

Descriptors—Achievement Tests, Aptitude Tests, Bulletins, *Early Childhood, *Measurement Instruments, Psychological Tests, *Publications, Publishing Industry, *Tests

Identifiers—Project Head Start

This bulletin lists tests and other measurement devices. It also includes publishers' catalogs and descriptive materials, information on scoring services and systems, test reviews, and reference materials on measurement and evaluation. Bulletin entries of particular interest to those working with children aged from birth to nine are preceded by a special symbol. The acquisitions are listed under the following headings: Achievement, Aptitude, Personality, Interests, Attitudes,

and Opinions, Miscellaneous, Sensory-Motor, Unidentified Announcements of publications, tests no longer available, test reviews, new references, new publishers, new addresses, addresses of publishers and organizations, an index, and subscription information are provided. (DB)

ED 073 168 TM 002 418

Ebel, Robert L.
Essentials of Educational Measurement.

Pub Date 72

Note—634p

Available from—Prentice-Hall, Inc., Englewood Cliffs, New Jersey (no price quoted)

Document Not Available from EDRS.

Descriptors—Achievement Tests, Aptitude Tests, Attitude Tests, Educational History, Educational Philosophy, *Educational Testing, Essay Tests, *Higher Education, Intelligence Tests, Interest Tests, Item Analysis, *Measurement, Multiple Choice Tests, Norms, Personality Tests, Scores, Standardized Tests, Test Construction, Testing, Testing Programs, Test Interpretation, Test Reliability, Test Validity, *Textbooks

This textbook is intended for use in an introductory course in the use of tests in schools and colleges. The book is divided into five parts: (1) history and philosophy—educational measurement, historical perspectives, Measurement and the process of education; (2) classroom test development—what should achievement tests measure?; how should achievement be measured?; how to plan a classroom test; the characteristics and uses of essay tests, true/false test items; how to write multiple-choice test items; (3) getting, interpreting, and using test scores—how to administer and score an achievement test, oral examinations, test score statistics, marks and marking systems; (4) test analysis and evaluation—how to judge the quality of a classroom test, how to improve test quality through item analysis, how to estimate, interpret and improve test reliability; the validity of classroom tests; and (5) published tests and testing programs—the scope and use of published tests, standardized achievement tests and test batteries, standard scores, norms, and the passing score, intelligence and aptitude tests, personality, attitudes, interests, and testing programs. A glossary of terms used in educational measurement, projects and problems, and a bibliography are presented in appendices. (KM)

ED 073 776 CG 007 784

Hedl, John J., Jr.
Test Anxiety: A State or Trait Concept?
Pub Date 71

Note—10p

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Anxiety, *Psychological Patterns, *Psychological Studies, Research Projects, *Stress Variables, Students, Student Testing, *Testing, Testing Problems

The purpose of this experiment was to relate two bodies of research on anxiety: test anxiety, or anxiety proneness specific to the testing situation, and trait-state anxiety. The author hypothesized that state anxiety—anxiety not tied to any particular situation but aroused in "any" situation—should be highly related to test anxiety during a testing situation; and, on the other hand, that state anxiety aroused during a situation of minimal evaluative stress, such as a game, should be less closely related to test anxiety. A total of 60 students were tested in both situations: a game involving no explicit evaluative stress and administration of the Simon Intelligence Test via computer, a situation of some evaluative stress. Measures of state anxiety were obtained before and after each of these situations. The results failed to confirm the hypothesis. Since the initial analysis indicated that test anxiety was not as responsive to situational stress as initially hypothesized, a succeeding analysis was conducted to determine whether test anxiety was more similar to the construct of trait anxiety. The results of this study showed that the test anxiety construct is not as responsive to situational stress as is the state anxiety measure. References are included. (Author/SES)

ED 075 445

TM 002 522

Cawdrey, Thomas M., Jr., Comp.

Invitational Conference on Measurement in Education: Proceedings of the Eleventh Southeastern Conference (December 8-9, 1972). Pre-Session.

Georgia Univ., Athens Coll. of Education
Spons Agency—Harcourt Brace Jovanovich, Inc.,
New York, N.Y. Test Dept.
Pub Date Dec 72

Note—45p. Papers presented at pre-session of Southeastern Invitational Conference on Measurement in Education (11th, Athens, Georgia, December 8, 1972).

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Conference Reports, Educational Needs, Educational Planning, Educational Policy, Educational Research, Evaluation, Measurement, Speeches, Teacher Education.

Three papers presented at the pre-session to the 11th Invitational Conference on Measurement in Education are provided. These papers are: "Public Schools in Transition: One Student Body, Its Needs and Directions" by George R. Rhodes, Jr.; "Man to Mankind: The International Dimension of Teacher Education" by David G. Imig; and "Measurement for the Purpose of Evaluation" by D. E. Hinkle. (For related documents, see TM 002 523-529, 541-547. (DH))

ED 075 449

TM 002 526

Cawdrey, Thomas M., Jr., Comp.

Invitational Conference on Measurement in Education: Proceedings of the Eleventh Southeastern Conference (December 8-9, 1972). Session I.

Georgia Univ., Athens Coll. of Education
Spons Agency—Harcourt Brace Jovanovich, Inc.,
New York, N.Y. Test Dept.
Pub Date 8 Dec 72

Note—211p. Papers presented at Session I of Southeastern Invitational Conference on Measurement in Education (11th, Athens, Georgia, December 8, 1972).

EDRS Price MF-\$0.65 HC-\$9.87

Descriptors—American Indians, Conference Reports, Disadvantaged Youth, Educational Research, Measurement, Mental Retardation, Negro Students, School Integration, Special Education, Speeches.

Eleven papers presented at Session I of the 11th Invitational Conference on Measurement in Education are provided. This session focuses on special education subgroups, especially the disadvantaged. Particular emphasis was placed on mental retardation, American Indian education, school desegregation, and the gifted disadvantaged black. The eleven papers are: "Hazards in Research Involving Minorities" by Junius A. Davis; "The Self Concept of Mental Retardates: Some Problems of Assessment" by John I. Shultz; "Non-Cognitive Development of Mentally Retarded Children" by Bert D. Richmond; "Assessment of Motor Learning Through Mathematical Analysis" by Ernest I. Bandischoff and Susan J. Gosh; "Evaluating Achievement of the Mentally Retarded: A Comprehensive Process" by Andrea I. Shostick; "Measurement and Testing Considerations for Native American Education" by Joseph D. Blanchard; "School Desegregation: The Problem" by Milton Hill; "Perspectives of School Desegregation in the Southeast" by Merrill M. Hall and Harry B. Williams; "Identification of Gifted and Creative Children and Youth Among Black Disadvantaged Groups" a symposium; "The Prediction of Achievement Means of Schools from Non-School Factors Through Criterion Scaling" by Tom C. Innes; and "Information Feed-Back Systems and Educational Goals" by W. A. F. White. (For related documents, see TM 002 522-525, 527-547.) (KM)

ED 075 488

TM 002 567

Nake, Robert E.

School Accountability Laws.

Pub Date 72

Note—4p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Educational Accountability, Educational Legislation, Guides, School Responsibility, State Legislation.

An accountable school is seen as being one that (1) discloses its activities, (2) makes good

on staff promises, (3) assigns staff responsibility for each area of public concern, and (4) monitors its teaching and learning. School accountability laws are seen as a national demand that must be satisfied. To satisfy the present demands for legislation, to help the schools change, and to be least damaging to what the schools are doing well, a school accountability law should call for the following characteristics: (1) commitments from schools to provide high-quality teaching, opportunities for enriching experiences, and accommodation to the individual needs and aspirations of students and teachers; (2) an annual visit to each school by at least one committee of citizens; (3) judgment as to quality of instruction made by professional educators both inside and outside the school; (4) standardized measures of student basic skills by testing and by professional observation; (5) updated specifications of the errors of measurement for individual student scores, group means, and differences of both, as well as errors of measurement in nonstandardized observations and professional judgment; (6) each school should have a confidentiality policy; (7) each school should have a grievance procedure; (8) one or more centers should be established for the evaluation of individual teacher competence; and (9) schools should be supported in efforts to keep bureaucracy to a minimum. To be in the best interests of the people, state accountability laws should protect local control of the school, individuality of teachers, and diversity of learning opportunities. (DH)

ED 075 541

UD 013 507

Shanner, William M.

Procedures for the Preparation of Descriptions on Assessment of Exemplary Programs. Final Report.

American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
Spons Agency—Office of Education (DHEW),
Washington, D.C. Office of Information Dissemination.

Report No.—AIR 20200 4-73 FR

Bureau No.—BR 2-0200

Pub Date Apr 73

Contract—OEC-0-70-3674

Note—140p.

EDRS Price MF-\$0.65 HC-\$6.58

Descriptors—Basic Skills, Bilingual Education, Compensatory Education Programs, Educational Change, Elementary Education, Information Dissemination, Inner City, Instructional Materials, Program Descriptions, Program Effectiveness, Program Evaluation, Reading Programs, Secondary Education, Spanish Speaking, Vocational Education.

The chief purpose of this project was to prepare descriptions of exemplary programs in childhood education and reading having sufficient merit to be considered as visitation sites for other educators. The childhood education phase of this project was funded jointly by the Office of Economic Opportunity and the Office of Education. Its purpose was to select and document a variety of programs including Head Start, Follow Through, elementary and secondary programs in inner city schools, programs for Spanish Speaking, and programs in basic skills and vocational areas for initial dissemination through the six Regional and National White House Conferences. In addition it was planned that OEO and OE could use the print and nonprint materials developed by the project for independent dissemination efforts. The exemplary reading programs phase of the project involved the field assessment of selected programs related to reading instruction and the verification of their effectiveness as demonstration sites for visits by personnel from other school districts. Program descriptions and leaflets were prepared for the 37 reading programs and 34 childhood education programs. Audio-visual documentation was prepared for 12 childhood education programs. This project was one in a series directed toward the wider dissemination of new and innovative practices in education, and was funded by the Office of Information Dissemination of OE. (Author/IM)

59 076 452

50 005 529

Elliott, Richard J.

Re-examining the Notion of the Universality of

Educational Measurement.

Pub Date 22 Nov 72

Note—12p. Paper presented at the Annual Meeting of the National Council for the Social Studies, Boston, November, 1972.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Attitudes, Educational Testing, Evaluation Criteria, Interaction, Measurement, Objectives, Personal Values, Social Sciences, Social Values, Speeches, Testing, Values.

Identifiers—National Council for the Social Studies.

The purpose of this paper is to critically raise some questions as to whether measurement can provide us with a universally applicable set of criteria for making value judgments in the social sciences. Stated are assumptions that provide the framework of current standardized testing. Questions are raised as to the efficacy of current standardized testing in the social sciences because of the one-dimensional aspect of measurements. Four assumptions about education, experience, reality, and value formation are identified to provide, hopefully, a fuller reference point for social science educators when viewing current standardized measurement practices within the social sciences. The central theme of the paper suggests that, if these four assumptions hold true, they greatly limit the possibility of measurement providing a universally applicable set of criteria for making value judgments in the social sciences. (Author)

ED 076 702

TM 002 711

Ebel, Robert L.

The Future of Measurements of Abilities II.

Pub Date 27 Feb 73

Note—22p. Speech given before the annual meeting of American Educational Research Association (New Orleans, Louisiana, February 25-March 1, 1973).

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Achievement Tests, Aptitude Tests, Educational Testing, Intelligence Tests, Measurement Goals, Measurement Instruments, Speeches, Student Evaluation, Test Interpretation, Test Validity.

Identifiers—Thorndike (E. L.)

The views of E. L. Thorndike on the future of measurements of abilities, expressed 25 years ago, are summarized, and the future of measurements of abilities as it appears now is examined. Opportunities for improvement now arise mainly from increasing social concern for effective education. Measurement technology has developed rapidly and cannot continue at the same pace. Newer instructional technologies will not be the most widely used because they are costly, impersonal, inflexible, and less learner-oriented. Formative evaluation can supplement but not replace summative evaluation, and criterion-referenced testing can supplement but not replace norm-referenced testing. The concept of mastery learning cannot be applied rigorously to most tests of abilities, the learning of any complex skill of understanding is always incomplete. Social concern is evidenced in public demand for accountability and governmental desire to allocate funds more equitably. A serious problem in the use of tests of abilities is what to measure—what the proper roles are for intelligence tests, tests of general mental abilities, critical thinking tests, tests of creativity, and tests for affective outcomes. Some problems in determining test validity are created by asking the wrong questions and by not recognizing that each different test measures a somewhat different ability. The term "construct validity" is used loosely and with a variety of meanings. Employment tests have been questioned by courts recently, and their validity must be proved. Another major problem is that of developing the necessary tests in quantity and at low cost. (KM)

ED 077 933

TM 002 728

Weiss, David J., Britt, Nancy F.

Ability Measurement: Conventional or Adaptive? Minnesota Univ., Minneapolis: Dept. of Psychology.

Spons Agency—Office of Naval Research, Washington, D.C. Personnel and Training Research Programs Office.
Report No.—RR-73-1

Pub Date Feb 73

Note 77p

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors: Ability, Comparative Analysis, Group Tests, Individual Tests, Literature Reviews, Measurement Instruments, Psychometrics, Testing, Test Reliability, Test Validity

Research to date on adaptive (sequential, branched, individualized, tailored, programmed, response contingent) ability testing is reviewed and summarized, following a brief review of problems inherent in conventional individual and group approaches to ability measurement. Research reviewed includes empirical, simulation and theoretical studies of adaptive testing strategies. Adaptive strategies identified in the literature include two-stage testing and multistage tests. Multistage tests are differentiated into fixed branching models and variable branching models (including Bayesian and non-Bayesian strategies). Results of research using the various strategies and research approaches are compared and summarized, leading to the general conclusion that adaptive testing can considerably reduce testing time and at the same time yield scores of higher reliability and validity than conventional tests, under a number of circumstances. A number of new psychometric problems raised by adaptive testing are discussed, as is the criterion problem in evaluating the utility of adaptive testing. Problems of implementing adaptive testing in a paper and pencil mode, or using special testing machines are reviewed, the advantages of computer-controlled adaptive test administration are described. (Author)

ED 077 964

TM 002 773

Reilly, Richard R.

A Note on Minority Group Test Bias Studies.

Educational Testing Service, Princeton, N.J.

Report No. ETS RB 73-15

Pub Date Feb 73

Note 11p

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors: Comparative Analysis, Minority Groups, Prediction, Statistical Analysis, Technical Reports, Test Bias, Test Interpretation

Comparisons of majority and minority group regression lines for purposes of assessing test bias may, under certain conditions, be viewed as comparisons of conditional bivariate distributions. Where these conditions hold, findings should reveal parallel regression lines except for a special case. Given the conditions described, one implication is that even when the test is a parallel form of the criterion, lines with equal slopes but unequal intercepts should be found. (Author)

ED 077 995

TM 002 807

Hauscher, Ronald E.

Some Points of Confusion in Discussing the Testing of Black Students.

Educational Testing Service, Princeton, N.J.

Report No. ETS RM 73-5

Pub Date Mar 73

Note 10p. Paper prepared for a Symposium of the American Educational Research Association Annual Meeting (New Orleans, Louisiana, February 26, 1973).

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors: Communication Problems, Educational Testing, Negro Students, Psychometrics, Racial Discrimination, Speeches, Standardized Tests, Test Bias, Test Interpretation, Test Validity

Four confusing issues that have delayed progress toward an awareness that testing is not a source of unfairness for minority students are discussed: (1) the assumptions underlying most of our psychometric manipulations are often not acknowledged or understood; (2) the extent of the objectivity of psychometrics is frequently exaggerated; (3) the meaning of certain terms, particularly "validity" (largely because it has both a technical and common usage), is quite confused; and (4) the understanding of just what function the tests are serving shifts from one function to another, unnoticed by those concerned. (KM)

ED 078 006

TM 002 818

Burns, Edward

IQ Testing and Minority School Children: Imperatives for Change.

Connecticut Univ., Storrs. National Leadership Inst. Teacher Education/Early Childhood Note 8p

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors: Ability Identification, Culture Free Tests, Intelligence Quotient, Intelligence Tests, Minority Group Children, Performance Factors, Publications, Student Ability, Test Bias, Test Interpretation

The inadequacy and misuse of intelligence testing for minority group children are examined. IQ test items, norms, examining procedures, and language usage are discussed in terms of their bias against minority children. The implications of this bias for the classroom teacher are explored with the view that teacher mental sets are powerful mediators in learning and performance and that intelligence test scores play a major role in determining the nature of the set teachers develop. Culture specific or environment testing and criterion referenced tests are discussed as nondiscriminatory and more valuable methods of evaluation. (KM)

ED 078 020

TM 002 835

Popham, W. James. And Others.

Of Measurement and Mistakes.

Pub Date 29 Mar 73

Note 6p. Testimony before the General Subcommittee on Education, Committee on Education and Labor, U.S. House of Representatives, Washington, D.C., March 29, 1973.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors: Comparative Analysis, Criterion Referenced Tests, Evaluation Techniques, Measurement Goals, Measurement Techniques, Norm Referenced Tests, Program Evaluation, Research Methodology, Speeches, Student Evaluation

Because of misconceptions regarding appropriate measurement strategies, it is necessary to draw distinctions between two major measurement methodologies: norm referenced and criterion-referenced measurement, as they relate to determining basic academic capabilities. Norm-referenced measures are used to ascertain an individual's performance in relationship to the performance of other individuals on the same measuring device. Criterion-referenced measures are used to ascertain an individual's status with respect to some criterion, that is, an explicitly described type of learner competence. Because of the wide use of norm-referenced standardized achievement tests, many assume that they are the only instruments that should be used to find out how well a school is working or a pupil is learning. But typical standardized tests are unsuitable for these purposes because of problems with their interpretability and their psychometric properties. Criterion-referenced tests remedy some of these weaknesses because they can: (1) be more accurately interpretable; (2) detect the effects of good instruction; and (3) allow us to make more accurate diagnoses of individual learners' capabilities. If sufficient care is taken to support the development of high quality criterion-referenced measures, legislation to distribute federal funds on the basis of educational deficiencies rather than census determinants appears to be sound. (Author/KM)

ED 078 032

TM 002 847

Does Testing Cost Too Much?

Psychological Corp., New York, N.Y.

Pub Date Sep 71

Note 2p. Reprint.

Available from: Not available separately, see TM 002 846.

Journal Cit.—Test Service Bulletin, n42 p10-11 Sep 1971.

Document Not Available from EDRS.

Descriptors: Bulletins, Counseling Effectiveness, Educational Testing, Estimated Costs, Expenditure Per Student, Guidance Counseling, Personnel Selection, Testing, Vocational Aptitude

Before a business adopts tests in personnel selection, it should be decided that the increased cost and bother are likely to be offset by the savings that come from this additional selection procedure. Fairly accurate estimates of the cost of testing can be made, but in educational testing it is harder than in businesses to measure the

results in dollars and cents. Most of the benefits appear in terms of pupil adjustment or in the student's post-graduate life. But in the absence of a capable counseling staff, the use of any tests at all may be an extravagance. Conversely, if a school is committed to a counseling service for its pupils, it is unwise not to provide the low-cost tools that can save the time of high-cost counselors. (For related documents, see TM 002 846, 848.) (KM)

ED 078 051

TM 002 866

Prediger, Dale J.

The Vital Role of Testing in Career Guidance.

Pub Date 14 Jan 73

Note—23p. Paper presented at a symposium of the APGA National Convention (San Diego, Calif., Feb. 1973).

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors: Career Planning, Counselors, Decision Making, Guidance Counseling, Occupational Guidance, Role Perception, Self Evaluation, Testing, Test Interpretation, Test Results, Theories

An attempt is made to show that career development theory and career decision-making theory both provide a firm foundation for the role of tests in career guidance. This role is primarily to stimulate, broaden, and provide focus to the exploration of self in relation to career. To effectively implement this role, test publishers must devote much more attention to transforming test scores (data) into counseling information. It is suggested that more than information is needed to obtain real benefit from tests. Counselors must help students transform the information obtained from tests into exploratory activities and self-evaluated experiences. For this reason, test interpretation can no longer be treated as a hot-house experience devoid of the real world. Career development and decision-making principles indicate that tests must be used in the context of a developmental career guidance program. When this is done, tests can play a vital role in career guidance. (Author/CK)

ED 078 316

CG 007 214

Lewin, Roger T.

Testing and Equality: The Great Divorce?

Pub Date Feb 72

Note—9p. Paper presented at the California Personnel and Guidance Association convention in Los Angeles, California on February 19-21, 1972.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors: Disadvantaged Groups, Disadvantaged Youth, Equal Facilities, Equal Opportunities (Jobs), Minority Groups, Norm Referenced Tests, Speeches, Standardized Tests, Student Testing, Test Bias, Testing, Testing Problems, Tests

The paper examines the issue of whether the concepts and practices of standardized testing, on the one hand, and the pursuit and attainment of equal opportunity, on the other, are compatible. The author cites a number of areas where tension and conflict have centered, however, he feels that there is no necessary incompatibility between the assessment of human talents and the pursuit of both equality and excellence. A number of suggestions are discussed which might help to ameliorate some of the problems of testing of minority groups: 1) keep the discussion at a rational, rather than inflammatory, level; 2) explore new instrument development to more accurately capture the magnitude and variety of abilities of minorities; 3) expand training in the administration and interpretation of present tests with minority examinees; 4) adopt more systematic procedures for review and analysis of tests; 5) cross-check and verify test scores against other signs of talent; and 6) use tests humanely. (Author/SES)

ED 079 333

TM 002 931

Ayabe, Harold E.

Measuring Reflection Impulsivity Accurately.

Note—14p. Paper presented at annual meeting of American Educational Research Association (New Orleans, Louisiana, February 25-March 1, 1973).

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Age Differences, *Conceptual Tempo, Correlation, Kindergarten, *Measurement Instruments, Primary Grades, *Psychometrics, Speeches, *Student Testing, Test Results

Identifiers—Matching Familiar Figures Test, Multi-Choice Conceptual Tempo Apparatus

Two hundred children (100 girls and 100 boys), kindergarten through fourth grade, were tested with the Matching Familiar Figures Test (MFF) and a newer instrument, the Multi-Choice Conceptual Tempo Apparatus (MCCTA). The MCCTA was more reliable and had stronger negative correlations between errors and latency than did the MFF. The MCCTA did not correlate with either School and College Ability Test or the California Achievement Test on verbal, numerical and total scores. MCCTA errors correlated positively with MFF errors. High latency scores on the MCCTA tended to also delay their response on the MFF. The MCCTA appears to be a better instrument for the measurement of Conceptual Tempo than the MFF. (Author)

ED 079 345

TM 002 944

Doppelt, Jerome E.

How Accurate Is a Test Score?

Psychological Corp., New York, N.Y.

Report No.—Bull-50

Pub. Date Jun 56

Note—3p., Reprint from Test Service Bulletin.

Journal Cit—Test Service Bulletin, n50 p1-3 Jun 1956.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Bulletins, *Measurement Techniques, Reliability, *Scoring Formulas, *Standard Error of Measurement, Statistical Analysis, *Test Results, *True Scores

The standard error of measurement as a means for estimating the margin of error that should be allowed for in test scores is discussed. The true score measures the performance that is characteristic of the person tested, the variations, plus and minus, around the true score describe a characteristic of the test. When the standard deviation is used as a measure of the variation of observed scores around the true score, the result is called the standard error of measurement. The standard error of measurement can be used in defining limits around the observed score within which one would be reasonably sure to find the true score. Since, in practice, it is not possible to give a large number of equivalent forms of a test in order to find the characteristic standard error of measurement, it is determined by the reliability coefficient. As measured by the reliability coefficient, reliability means consistency of measurement. It is unfortunately true that a test will have different reliability coefficients depending on the groups of people tested. The standard error of measurement is less subject to this variation. The formula for computing it, which is given, takes into account both the reliability coefficient and the standard deviation for each group. A table is provided of Standard Errors of Measurement for Given Values of Reliability Coefficient and Standard Deviation. (For related document, see TM 002 943, 946.) (DH)

ED 079 431

TM 003 036

Instruments That Measure Self Concept.

Chicago Board of Education, Ill.

Note—81p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Bibliographies, Educational Quality,

*Elementary Grades, *Measurement Instruments, Program Evaluation, Psychological Tests, Rating Scales, *Self Concept, *Self Concept Tests, Urban Schools

Materials are presented that resulted from a search for a measure of self-concept to be used in the evaluation of Project RE.A.D., a program to improve the quality of education in urban schools. Most of the instruments listed are appropriate for the elementary level. In addition to a list of the instruments, this document also contains a listing of authors, a description of instruments, and a bibliography. (CK)

ED 080 544

TM 003 049

Developments and Assessments in Educational

Centers and Laboratories; The Annual Western Regional Conference on Testing Problems (10th, San Francisco, California, May 9, 1969).

Educational Testing Service, Princeton, N.J.

Pub. Date 9 May 69

Note—79p.

Available from Educational Testing Service, Princeton, New Jersey 08540 (\$2.00)

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Conference Reports, *Educational Development, Educational Research, Evaluation Criteria, Evaluation Methods, *Learning Laboratories, *Program Evaluation, *Testing Problems

The 1969 Western Regional Conference on Testing Problems dealt with developments and assessments in educational centers and laboratories. The following speeches were presented: (1) "Behavioral Objective Specifications in Evaluation: Relevant or Irrelevant?" by Marvin C. Alkin, (2) "Approaches to the Validation of Learning Hierarchies" by Margaret C. Wang, (3) "Some Problems with Regard to Research and Development in Higher Education" by Leland L. Medsker, (4) "Educational Research, Educational Development and Evaluation Studies" by John K. Hemphill, and (5) "The Challenge of Multi-Agency Involvement in Development" by Ray Jorgeward. A list of conference participants concludes the report. (KM)

ED 080 560

TM 003 065

Hanger, Ronald J.

Research on Testing and the Minority Student.

Educational Testing Service, Princeton, N.J.

Pub. Date 29 Nov 72.

Note—6p., Paper presented at the GRE Board Research Seminar of the Council of Graduate Schools Annual Meeting (12th, New Orleans, Louisiana, November 29, 1972).

Available from: Not available separately, see TM 003 064.

Document Not Available from EDRS.

Descriptors—Achievement Tests, Aptitude Tests, *Graduate Study, Item Analysis, *Minority Groups, Predictive Validity, Summaries, Speeches, *Standardized Tests, *Test Bias, Testing

Identifiers—*Graduate Record Examinations.

There are many potential sources of test bias besides that of the particular item content within the test. The other potential sources, designated here as program and utilization must also be encompassed in any thorough and effective program to increase the accuracy of assessment for members of ethnic minorities. As usual, the research findings are emerging much more slowly than we would like, but that is the nature of careful research. Meanwhile, our failure to find bias from those sources that are most often identified, such as test content or predictive validity, must not be used to justify an abandonment of the search. The research efforts must encompass these other possible sources of inaccuracy, too, for that matter, they should be continuous, serving a monitoring function of these possibilities. But meanwhile, we can be aware of them and of the things that can be done to increase measurement accuracy using research evidence or lacking that, just good judgment and sensitivity. (For related documents, see TM 003 063, 064, TM 002 964.) (Author/KM)

ED 080 566

TM 003 072

Changes in Education and Measurement Since

Sputnik. Annual Western Regional Conference on Testing Problems (10th, Los Angeles, California, May 5, 1961).

Educational Testing Service, Princeton, N.J.

Pub. Date 5 May 61

Note—90p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Conference Reports, *Educational Change, English Instruction, Language Instruction, Language Tests, Mathematics Instruction, *Measurement, Science Curriculum, Science Tests, *Testing Problems

The 1961 meeting of the Western Regional Conference on Testing Problems dealt with changes in education and measurement since Sputnik. The following papers were presented: (1) "Who's Testing Whom and for What?" by Daniel D. Feder, (2) "Recent Development and Problems in the Teaching of English" by Alfred H. Grommon, (3) "The New Foreign Language

Teaching and the Need for Tests in All the Skills" by George Scherer, (4) "The Teaching of Modern Mathematics" by John L. Kelley, and (5) "Some New Science Curricula and their Measurement" by Frederick L. Ferris, Jr. A list of conference participants concludes the report. (KM)

ED 080 590

TM 003 107

Schroth, Marvin I.

The Use of IQ as a Measure of Learning Rate with Minority Children. Final Report.

Santa Clara Univ., Calif.

Spons. Agency—National Inst. of Education (DHEW), Washington, D.C.

Bureau No.—HR-2-1021

Pub. Date May 73

Contract—OEC-0-72-0032(057)

Note—12p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Educational Research, Elementary Grades, *Intelligence Quotient, *Learning Processes, Learning Theories, *Measurement Instruments, *Mexican Americans, *Socioeconomic Status, Technical Reports

The purpose of this study was to investigate the use of IQ as a measure of learning rate with children of similar ethnic origin but different socioeconomic backgrounds. The theoretical basis for this experiment was Jensen and Rohrer's distinction between developmental rate and learning rate. Specifically, a dimension abstracted oddity task was presented to 180 Mexican-American children with different IQs from lower and middle socioeconomic backgrounds. They were matched on MA. The results support the Jensen and Rohrer hypothesis as IQ was found to determine the rate of learning the task. Also, middle-SES children learned the task faster than their low-SES peers. (Author)

ED 080 918

CG 008 125

Hazeur, Felice B.

Symposium on "Dealing With Types of Test Biases When Black Students Are Concerned"—The Transition from High School to College: How Do We Bridge the Gap?

Pub. Date 73

Note—12p., Paper presented at the American Personnel and Guidance Association Meeting, 9-12 February 1973, San Diego, California

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Adjustment Problems, *College Freshmen, *Cultural Differences, Culture Free Tests, Motivation, Negro Education, *Negro Students, Racial Factors, Symposia, *Test Bias, Urban Education

In determining whether black freshmen are ready for college, the author examines the social-psychological realities from which they emerge. A poor concept of efficient utilization of time and organization, coupled with a weak academic self-discipline perpetuates a dangerous beginning for the black freshman. Parental lack of exposure to college does little to reinforce motivation, sense of direction, and self-concept. The document discusses ways of coping with the black student's problems in higher education and defines separation as a failure to prepare him for the culture in which he must function. The author also condemns one-way integration as a source of personal maladjustment for many black students and as a reinforcer of the dichotomy in American life. She concludes that only when we accept and appreciate mutual peculiarities and differences will our efforts take on new meaning and become viable exercise of professional responsibility. (Author/LAA)

ED 081 781

TM 003 141

Katz, Martin

Selecting an Achievement Test: Principles and Procedures.

Educational Testing Service, Princeton, N.J.

Pub. Date 73

Note—12p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Achievement Tests, Educational Testing, *Norms, Standardized Tests, *Test Reliability, Test Results, Tests, *Test Selection.

*Test Validity

The recommended procedure for achievement test selection is presented. Following a review of some fundamental definitions of statistical properties of achievement tests (validity, reliability, and norms), major considerations to be applied in selecting achievement tests are discussed. These considerations are: (1) the characteristics of the school population, (2) the content and objectives of the curriculum, (3) the purposes of testing, and (4) the use of test scores. An appendix provided a list of U.S. publishers of standardized tests who issue test catalogs. (DB)

ED 081 803

TM 003 164

Hoepfner, Ralph

The Validity of Tests of Social Intelligence.

California Univ., Los Angeles. Center for the Study of Evaluation

Spons. Agency—Office of Education (DHEW), Washington, D.C.

Report No.—CSF-R-85

Pub Date Apr 73

Note—35p. Address presented to the German Congress of Psychology (Saarbrücken, Germany, October 1972).

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Behavioral Science, Research, Evaluation Techniques, Factor Analysis, Intelligence Tests, Models, Performance Tests, Predictive Validity, Psychological Tests, Social Psychology, Speeches, Test Construction, Tests, Test Validity

Identifiers—Social Intelligence

The rationale underlying tests of social intelligence and some of the problems inherent in those tests are discussed. To measure social intelligence, paper and pencil tests were developed which were "situation free." These tests employed "stereotypic behavior of individual others." The stimuli used in the behavioral tests were photographs of people, artists' drawings, and cartoons. The types of responses were generally limited to marking on an answer sheet or writing a response. A series of tests for behavioral creativity that did not use paper and pencil, answer sheets, or writing were developed. Four types of performance tests were developed and administered to a sample of 30 people along with the paper-and-pencil tests. A new set of tests, approximately 25 to 30, was developed for behavioral production. Conclusions reached as a result of two validation studies are that people with high verbal intelligence don't need behavioral intelligence to perform well on the tests, and individuals with lower intelligence to perform well on the tests, and individuals with lower intelligence sometimes still achieved very high behavioral scores. Studies have also been done relating social intelligence to certain demographic variables. A study investigating the prediction of academic achievement correlated the behavioral-cognition tests with grades at a junior college; correlations of about .50 were found with achievement in various course grades. Three studies of the use of behavioral-cognition tests for predicting vocational success are discussed. Portions of the tests relating to factors of behavioral intelligence and factors of social creativity and their tests are provided. (DB)

ED 081 839

TM 003 200

Summary Report of Statewide Testing Program 1971-1972. Evaluation Report No. 81.

Hawaii State Dept. of Education, Honolulu. Office of Instructional Services.

Report No.—Eval-R-81

Pub Date 72

Note—60p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Achievement Tests, Aptitude Tests, Norms, Standardized Tests, State Programs, Tables (Data), Technical Reports, Testing Programs, Test Results

Hawaii's testing program for 1971-72 is described, and test results are provided. Following an introduction and a glossary of technical terminology, narrative summaries of all the tests administered are provided. These tests are: California Test of Mental Maturity, California Achievement Test—Reading, SCAT, STEP—Reading, Mathematics, Writing, Science, Social Stu-

dies, and Listening, and Differential Aptitude Test. In each narrative, the following information is given: purpose of the test, population taking the test, date of the test administration, and a summary of results. Appendices provide specific results for each test and State and national norms (KM)

ED 081 843

TM 003 204

Smith, Charles W.

Criterion-Referenced Assessment.

Pub Date 17 Jul 73

Note—14p. Paper presented at International Symposium on Educational Testing (the Hague, The Netherlands, July 17, 1973).

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Achievement Tests, Criterion-Referenced Tests, Educational Testing, Measurement Techniques, Norm-Referenced Tests, Speeches, Standardized Tests, Test Interpretation, Test Selection

Identifiers—Mastery Learning

Both criterion-referenced and norm-referenced measures are useful tools to the classroom teacher, but each has its specific uses. The criterion-referenced measure is useful when one is interested in whether an individual possesses particular competencies and when there are no quotas as to how many possess that skill. It is particularly useful in assessing competence in licensed professions since tasks in these areas must be performed at specifically high levels of competence. Criterion-referenced assessment is also important to any subject area where future academic success is dependent upon cumulative information or skills, such as in mathematics. The norm-referenced measure should be used when selectivity is required, such as in choosing the most able candidate to fill a position or when only a limited number of candidates can be selected for vocational training or academic pursuit. The criterion-referenced measure points out whether an individual possesses particular skills or competencies, but the norm-referenced measure is better able to indicate how well the individual performs in his competent area. The criterion-referenced measure aims to discriminate between successive performance of a given individual, while the norm-referenced measure aims to discriminate between individuals within a particular group on a given measure. Criterion-referenced assessment, along with feedback and remedial procedures, can help teachers realize the goals of mastery learning with their students. (Author/KM)

ED 082 061

CE 000 371

Fadale, LaVerna M.

An Instrument to Determine Career Awareness.

State Univ. of New York, Ithaca. Cornell Inst. for Research and Development in Occupational Education.

Spons. Agency—New York State Education Dept., Albany. Office of Occupational Education.

Report No.—Res-Pub-74-1

Pub Date Aug 73

Note—40p. For Student Booklet, see CE 000372

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Administrator Guides, Career Education, Educational Research, Elementary Grades, Elementary School Curriculum, Measurement Instruments, Models, Occupational Clusters, Test Construction, Test Results

Identifiers—Career Awareness

The project focused on the awareness role of the elementary school within career education and is based firmly on the School-board Comprehensive Career Education Model and the Occupational Clustering System. A theoretical basis for the element of career awareness in career education was defined. The existence of career awareness and the differences within this element were determined by an original instrument designed and validated for use at these age levels. Implications for educators and for curricular implementation were suggested by the research findings. The instrument is intended as a tool for classroom teachers and other educators interested in assessing the career awareness element as demonstrated by their students. An administration manual, sample instrument, and two-page bibliography are included. (Author/AG)

ED 082 062

CE 000 372

Fadale, LaVerna M.

Career Awareness Inventory. Student Booklet.

State Univ. of New York, Ithaca. Cornell Inst. for Research and Development in Occupational Education.

Spons. Agency—New York State Education Dept., Albany. Office of Occupational Education.

Pub Date 73

Note—27p. For Administrator and Research Guide to the Inventory, see CE 000 371

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Career Education, Educational Research, Elementary Grades, Measurement Instruments, Tests

Identifiers—Career Awareness

The Inventory was constructed to measure career awareness among elementary school students. In Test I students are shown a number of pictures portraying workers. They are asked to identify the worker's occupation in each situation. In Test II they identify which occupation requires a college education. Test III assesses the workers they may know. Test IV assesses their knowledge of workers who perform services from those who make products. Test V examines prestige; VI at occupational clusters; and VII at what workers in different jobs like to do. For information on its construction, administration, and interpretation, see CE 000 371. (AG)

ED 082 422

EC 060 093

Evaluation Bibliography.

North Carolina Univ., Chapel Hill. Technical Assistance Development System.

Spons. Agency—Office of Education (DHEW), Washington, D.C.

Pub Date 73

Note—41p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Annotated Bibliographies, Early Childhood Education, Exceptional Child Education, Handicapped Children, Parent Attitudes, Screening Tests, Testing

Identifiers—TADS, Technical Assistance Development System

Presented are brief descriptions of approximately 140 tests suitable for children under 4 years of age and 10 measures for use with parents. Tests for children are indexed by age range in months, by skills tested (language, cognition, self help, social-affective, visual motor, and physical health), and whether the test is designed for use in screening programs or to be completed by the parent. Usually given for a test is title, author, publisher, and a description covering test purposes, types of items, training needed to administer test, and reliability and validity studies. Examples of test for children described are Bayley Scales of Infant Development, Bender Motor Gestalt Test, Burke Behavior Rating Scales, and the Goldman-Fristoe Test of Articulation. Examples of tests for parents are the Baughman and Dahlstrom Parent Interview Schedule, Inventory of Home Stimulation, and the Porter Parental Acceptance Scale. Addresses of test publishers are included. (DB)

ED 083 308

TM 003 273

Vroom, Victor H.

A Multi-Dimensional Measure of Leader Behavior. Yale Univ., New Haven, Conn. Dept. of Administrative Science.

Spons. Agency—Office of Naval Research, Arlington, Va.

Report No.—AD-763-441; TR-4

Pub Date 21 May 73

Note—25p.

Available from: National Technical Information Service, Springfield, Va. 22151 (AD-763 441, MF \$1.45 HC \$3.00)

Document Not Available from EDRS.

Descriptors—Behavior Rating Scales, Factor Analysis, Leadership, Mathematical Models, Measurement Instruments, Technical Reports, Tests

The usefulness of the sets of cases developed in a previous technical report are examined as potential tests of leadership. Each set of cases generates, from a single subject, a number of scores which can be logically derived and which have substantial face validity. The intercorrelations among the scores, their split-half reliability,

and the correlations of scores with other measures of leadership behavior are examined, and conclusions are drawn concerning the use of such tests in leadership assessment. (Author)

ED 083 310 TM 003 275

Jackson, Steve. *Klinger, Ron*
Cross Cultural Attitude Inventory [and] Test Manual.
Education Service Center Region 13, Austin, Tex.
Spons. Agency—Bureau of Elementary and Secondary Education (DHEW/OE), Washington, D.C. Div of Bilingual Education
Note—64p

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Anglo Americans, *Attitude Tests, *Cultural Factors, Elementary School Students, Ethnic Groups, *Language Role, Manuals, Mexican Americans, Preschool Tests, Scoring, Secondary School Students, *Student Testing, *Test Construction, Tests
Identifiers—*Cross Cultural Attitude Inventory

The Cross Cultural Attitude Inventory provides a means for measuring the degree of positive or negative feeling which Mexican American and Anglo students (ages 3 to 18) have for the two cultures. Present test validity is based on face validity and the rationale behind the item selection and construction procedures. The inventory also provides a springboard from which other ideas may come. The inventory is comprised of 24 test items that relate to language, facial characteristics, foods, games, clothing, sports, and flags. The use of five faces beneath each item seems to work well with children in a wide range of age, sex, and ethnic categories. The test provides a score based on an equal number of items representative of each culture. It is a relatively quick instrument to administer (20-30 minutes), and is easy for both test administrator and subjects to understand. The age range of subjects who can be expected to understand and relate to the test is from 3 to 12 years, although older subjects have been used, the test should be administered individually to those of junior-high age or older. Administration of the test is discussed as to time considerations, physical facilities, materials, language considerations, special considerations by age group, and instructions to students (ages 3 through 8 and ages 9 and above) and terms to be used for test items. The test is scored by assigning a value of 1 to 5. A copy of the Test Data Recording Form is provided. (Author/DB)

ED 083 313 TM 003 305

Rosen, Pamela, Ed.
Test Collection Bulletin, Vol. 7, No. 4.
Educational Testing Service, Princeton, N.J.
Report No.—ETS Vol. 7, No. 4
Pub Date Oct 73
Note—22p

Available from—Test Collection, Educational Testing Service, Princeton, New Jersey 08540 (subscription \$2.00, \$2.50 foreign)

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Achievement Tests, Aptitude Tests, Attitude Tests, Bulletins, Personality Tests, Preschool Education, *Preschool Tests, Primary Grades, Publications, *Reference Materials, *Resource Guides, *Tests

Identifiers—Project Head Start

In this Test Collection Bulletin, annotated listings are provided of tests that have been acquired by Educational Testing Service, of tests that have been announced, of test reviews, and of new references. The test acquisitions are categorized as to type of test, as follows: Achievement, Aptitude, Personality Interests, Attitudes, and Opinions, and Miscellaneous, Sensory-motor, Unidentified. Tests that are included in a special Head Start Test Collection are identified with a symbol. In addition, tests that are no longer available are listed, scoring services and systems are cited, notes relating to tests and testing programs are provided, addresses of publishers and organizations whose materials and services are mentioned in this bulletin are given, and an index to Volume 7 is included. The final page of the bulletin provides subscription information relative to the Test Collection Bulletin. (DB)

ED 083 318 24 TM 003 319

Rosen, Pamela, Ed.
Tests For Educationally Disadvantaged Adults.
Educational Testing Service, Princeton, N.J. Test Collection

Pub Date Jul 73

Note—12p

Available from—Test Collection, Educational Testing Service, Rosedale Road, Princeton, N.J. 08540

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Achievement Tests, *Adult Basic Education, Adults, *Annotated Bibliographies, Aptitude Tests, *Dropouts, *Educationally Disadvantaged, English (Second Language), *Tests

Sixty-five instruments, published between 1925 and 1972, are described in this annotated bibliography. The devices are intended for adults who have received only an elementary education, and adults who have completed high school but whose education was impaired due to learning disabilities or other educational handicaps. Both achievement and aptitude measures are included, covering such areas as intelligence, ability, learning skills, non-verbal reasoning, vocabulary, reading, and mathematics. The Spanish editions of several tests in English as a second language are presented. The publisher's name and address is provided for each instrument. (NE)

ED 083 319 TM 003 320

Rosen, Pamela, Ed.
Self-Concept Measures; Grade 7 and Above.
Educational Testing Service, Princeton, N.J. Test Collection

Pub Date Jun 73

Note—7p

Available from—Test Collection, Educational Testing Service, Rosedale Road, Princeton, N.J. 08540

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Annotated Bibliographies, *Measurement Techniques, Personality Tests, *Secondary School Students, *Self Concept Tests, Self Esteem, *Student Attitudes

This 34-item annotated test bibliography deals with a variety of currently available measures of self-concept and self-esteem. For the purposes of this listing, self-concept was defined as a multidimensional construct encompassing the range of an individual's perceptions and evaluations of himself. Many of the devices contained herein emphasize the learner's self-concept or the individual's conceptions of himself in the school environment. However, several global measures are also described. Various methods for assessing self-concept, including direct observations, behavior ratings, self-reports, and projective techniques, are presented. The instruments described in this listing are appropriate for use in grade seven and above. Information was obtained from the holdings and references of the Educational Testing Service Test Collection. (Author)

ED 083 320 TM 003 321

Rosen, Pamela, Ed.
Measures of Self-Concept Grades 4-6.
Educational Testing Service, Princeton, N.J. Test Collection

Pub Date Feb 73

Note—6p

Available from—Test Collection, Educational Testing Service, Rosedale Road, Princeton, N.J. 08540

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Annotated Bibliographies, *Elementary School Students, *Measurement Techniques, *Self Concept Tests, Self Esteem, *Student Attitudes

This 31-item test bibliography deals with a variety of currently available measures of self-concept and self-esteem. For the purposes of this listing, self-concept was defined as a multidimensional construct encompassing the range of an individual's perceptions and evaluations of himself. Many of the devices contained herein emphasize the learner's self-concept or the child's conception of himself in the school environment. However, several global measures are also described. Various methods for assessing self-concept in-

cluding direct observations, behavior ratings, self-reports, and projective techniques, are presented. The instruments described in this listing are appropriate for use with children in grades four through six. Information was obtained from the holdings and references of the Educational Testing Service Test Collection. (Author)

ED 083 321 TM 003 322

Rosen, Pamela, Ed.
Attitudes Toward School and School Adjustment Grades 4-6.
Educational Testing Service, Princeton, N.J. Test Collection

Pub Date Feb 73

Note—8p

Available from—Test Collection, Educational Testing Service, Rosedale Road, Princeton, N.J. 08540

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Annotated Bibliographies, Elementary School Students, Interest Tests, *Measurement Techniques, School Environment, *Student Adjustment, *Student Attitudes, Student Interests, *Student School Relationship, Student Teacher Relationship

This 31-item test bibliography lists currently available measures of attitudes toward school and school adjustment. The construct—attitudes toward school—encompasses pupils' attitudes toward themselves as learners, learning as a process, the school environment or classroom situation, specific school subjects, and teachers. In addition, the pupils' behavior is considered if it is indicative of their adjustment or lack of adjustment to the educational environment. Teacher ratings, self-report devices, and observation techniques are the various methods for assessing these attitudinal elements which have been included in the listing. Instruments described in this bibliography are appropriate for use with students in grades four through six. Information was obtained from the holdings and references of the Educational Testing Service Test Collection. (Author)

ED 083 323 TM 003 324

Rosen, Pamela, Ed.
Attitudes Toward School and School Adjustment; Grades 7-12.
Educational Testing Service, Princeton, N.J. Test Collection

Pub Date Mar 73

Note—7p

Available from—Test Collection, Educational Testing Service, Rosedale Road, Princeton, N.J. 08540

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Annotated Bibliographies, Interest Tests, *Measurement Techniques, School Environment, Secondary School Students, *Student Adjustment, *Student Attitudes, Student Interests, *Student School Relationship, Student Teacher Relationship

This 53-item test bibliography lists currently available measures of attitudes toward school and school adjustment. The construct—attitudes toward school—encompasses pupils' attitudes toward themselves as learners, learning as a process, the school environment or classroom situation, specific school subject, and teachers. In addition, the pupils' behavior is considered if it is indicative of their adjustment or lack of adjustment to the educational environment. Teacher ratings, self-report devices, and observation techniques are the various methods for assessing these attitudinal elements which have been included in the listing. Instruments described in this bibliography are appropriate for use with students in grades seven through twelve. Information was obtained from the holdings and references of the Educational Testing Service Test Collection. (Author)

ED 083 325 UD 013 718

Barabas, Jon, Comp.
The Assessment of Minority Groups: An Annotated Bibliography. ERIC-IRCD Urban Dis-

advantaged Series, Number 34, August 1973.
Columbia Univ., New York, N.Y. ERIC
Clearinghouse on the Urban Disadvantaged
Rpns Agency—National Inst. of Education (D-
HEW), Washington, D.C.
Pub Date Aug 73
Contract—OEC-0-9-420088-3327
Note—85p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Achievement Tests, *Annotated
Bibliographies, Attitude Tests, Culture Free
Tests, *Educational Diagnosis, Intelligence
Tests, *Minority Groups, Personality Tests,
Predictive Ability (Testing), Racial Differences,
Social Differences, *Testing, *Testing
Problems, Test Reliability, Test Validity

The materials cited here represent information
on such diverse but interrelated areas as methods
of assessing achievement, intelligence, personality
factors, and attitudes, effects of testing on self
concept and employment opportunities, predic-
tion of academic success, reliability and validity
of specific tests, criticism of the methods and use
of assessment, test construction, use of assess-
ment for educational placement and diagnosis,
culture free and culture fair tests, performance
differences on tests between majority and minority
groups. The educational literature from the
early 1960's to the present from the following
sources was searched: the ERIC system, the
Library of Teachers College, Columbia University,
the in-house document collection of the ERIC
Information Retrieval Center on the Disad-
vantaged, and several existing bibliographies. A
few pieces dated prior to 1960 were included.
The bibliography is arranged alphabetically by
author and is followed by a subject index. Each
citation is cross-referenced under two or more of
the index terms which, hopefully, best describe
the material. The bibliography is completely an-
notated. For those documents which deal only in
part or peripherally with minority groups, relevant
quotations are given or the relevant sections
are described. A fuller annotation is included for
those materials which deal totally with minority
group assessment. (Author/JM)

ED 084 284 TM 003 392
Saarni, Carolyn L. And Others
The Vicissitudes of Sex-Role Assessment.
Pub Date Apr 73

Note—18p.; Paper presented at the Annual Meet-
ing of the Western Psychological Association
(Anaheim, Cal., April 1973)

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Analysis of Variance, *Identifica-
tion (Psychological), *Measurement Instru-
ments, *Psychological Studies, Rating Scales,
*Role Perception, Self Concept, *Sex Dif-
ferences, Tables (Data), Technical Reports
Identifiers—Psychological Femininity Scale, Sex
Role Stereotype Scale

A study of the psychological differences
between maleness and femaleness is presented.
The sample studied consisted of four groups: (1)
24 pre-service nursing trainees, all female; (2) 57
introductory psychology students, 31 females and
26 males; (3) 19 Caucasian non-college graduate
adults, 11 females and 8 males; and (4) 27 radical
feminists. Instruments used were Robert
May's projective test technique, the Sex Role
Stereotype Scale developed by Rosenkrantz,
Broverman et al., Gough's scale for assessing
"psychological femininity" and the process of
asking the subject what he thinks. A 4 X 3 X 2
multivariate analysis of variance for female sub-
jects and a 2 X 3 X 2 multivariate analysis of
variance for male subjects were conducted to test
for main effects and interactions of group mem-
bership, marital status, and having children or not
on the eight dependent variables derived from the
four sex role instruments. The results of the study
showed the following two consistencies: (1) males
receive substantially more masculine scores on
the Sex-Role Stereotype Scale and the Gough
Scale than women if the women's scores across
the four groups are combined; and (2) one's peer
group appears to influence the expression of sex-
typed behaviors and dispositions. As to the validity
of sex-role assessment, current tests and scales
do not distinguish among the differing degrees of
significance of various sex-typed behaviors but
rather sum or average across various items of sex-
typed behavior. (CK)

ED 084 285 TM 003 293

Green, Donald Ross
Racial and Ethnic Bias in Achievement Tests and
What To Do About It.

Note—9p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Achievement Tests, *Criterion
Referenced Tests, Cultural Differences, Ethnic
Groups, *Norm Referenced Tests, Problem
Solving, Program Evaluation, *Student Evaluation,
*Test Bias, Test Construction, Test
Results

A description of two proposals for alleviating
the racial and ethnic bias in tests of achievement
used in schools is presented. One of them entails
adding steps to the construction procedures used
in building norm referenced achievement tests;
the second entails using criterion-referenced
achievement tests rather than standardized tests
for certain purposes. The principal uses of
achievement tests are to: (1) evaluate the status
of a student or a set of students in a class, school,
or school system; (2) evaluate programs, curricula,
and instructional materials; (3) diagnose
problems; and (4) provide a basis for planning in-
dividual, class, or system programs. The bias built
into tests arises in the minds of those who write
and edit the tests and from the procedures used
to improve the tests. It is suggested that members
of each of the groups concerned with the test
participate in constructing the examinations from
the start and to use item writers and raters that
represent all major ethnic and cultural groups in
the population. Criterion-referenced tests should
be designed to show exactly what the pupils have
learned; these tests should be used for specific
diagnosis of school and program problems. (CK)

ED 084 488 CG 008 323

Parker, Paul J.
The Systematic Desensitization of High Debilitat-
ing Test Anxious College Students by Relaxation
and Assertion.

Pub Date Apr 73

Note—12p.; Paper presented at the Southwestern
Psychological Association Meeting, 4 April
1973, San Antonio, Texas

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Anxiety, Attitudes, Behavior,
*Behavior Change, College Students, *Desen-
sitization, *Progressive Relaxation,
Psychotherapy, *Testing Problems
Identifiers—*Achievement Anxiety Test

The present study compared the effects of as-
sertion with that of progressive relaxation training
in systematic desensitization. Nineteen Ss
were selected on the basis of exemplifying high de-
bilitating test anxiety according to Alpert and
Haber's (1960) Achievement Anxiety Test.
Results showed that test anxious Ss who received
either relaxation or assertive training experienced
a significantly greater reduction in debilitating
test anxiety on both the post- (p less than 0.001)
and six week follow-up (p less than 0.01) mea-
sures than no-treatment control Ss with cor-
responding pre-treatment scores. Although the
post-treatment measure indicated that relaxation
was significantly more effective than assertion,
the six week follow-up failed to reveal a signifi-
cant difference between the two parameters. In
addition, assertive training was shown to bring
about a significant reduction in test anxiety in al-
most half the Ss as relaxation. (Author)

ED 084 835 EM 011 656

Richardson, Frank C.
A Self-Study Manual For Students On Coping
With Test-Taking Anxiety.

Texas Univ., Austin. Computer-Assisted Instruc-
tion Lab.

Spons. Agency—National Science Foundation,
Washington, D.C.

Report No.—TR-25

Pub Date Sep 73

Note—83p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Anxiety, *Computer Assisted In-
struction, Independent Study, *Study Guides,
*Testing Problems, *Tests

A self-study manual for students on coping

with test-taking anxiety is presented along with a
commentary by its author. The manual is
designed for use in conjunction with videotapes
and practice of anxiety management techniques
in a computer-guided practice test-taking session.
The manual is part of a program designed to pro-
vide a regularly available service for highly test-
anxious students in computer-assisted instruction.
(RH)

ED 084 909 FL 004 533

Meeker, Mary Meeker, Robert
Strategies for Assessing Intellectual Patterns in
Black, Anglo, and Mexican-American Boys—or
Any Other Children—and Implications for Edu-
cation.

Pub Date [73]

Note—32p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Anglo Americans, Aptitude Tests,
Cognitive Tests, Cultural Factors, Culture Free
Tests, *Educational Testing, *Group Intelli-
gence Tests, Intelligence Quotient, *Intelli-
gence Tests, Mexican Americans, *Minority
Group Children, Negro Youth, Prognosis
Tests, Spanish Speaking, Student Testing, *Test
Bias, Test Construction, Test Interpretation,
Test Reliability, Test Validity

Identifiers—*Stanford Binet Intelligence Test

In this analysis of intelligence testing of minority
group children, the implications of inadequate
testing practices are discussed. Several aspects of
test design are examined: deficiencies in intelli-
gence testing, cultural bias, construct validity,
and diagnostic utility. A sample set of results
derived from a Stanford-Binet test administered to
257 respondents is examined; statistical data
are included. The author concludes that "in-
vestigations of cultural biases in intelligence test-
ing have established the fact that the most widely
used test procedures are 'penalizing' for non-Anglo,
lower socioeconomic groups." Teachers are
cautioned of the dangers in using group-test
results to plan academic programs geared to in-
dividual needs. (RL)

ED 085 404 TM 003 345

Rosen, Pamela, Ed.
Test Collection Bulletin Vol. 1, No. 3, April 1973.
Educational Testing Service, Princeton, N.J.
Pub Date Apr 73

Note—18p.

Available from—Test Collection, Educational
Testing Service, Princeton, NJ 08540 (Sub-
scription \$2.00, Foreign Readers \$2.50)

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Achievement Tests, Aptitude Tests,
Attitude Tests, *Bulletins, Elementary School
Students, Evaluation Techniques, Information
Dissemination, Interest Tests, *Measurement
Instruments, Personality Tests, *Preschool
Tests, Testing, Test Reviews, *Tests

Identifiers—*Head Start Test Collection

The Test Collection Bulletin is a quarterly
digest of information on tests and test related
services and activities. Each issue lists acquisitions
to the E-T-S-Test Collection, an extensive library
of tests, questionnaires, record and report forms
and other assessment devices. In this issue are
147 measures of achievement, aptitude, per-
sonality and attitudes, sensory-motor skills, and
miscellaneous skills. Also included are announce-
ments of tests, test reviews, new references on
measurement and evaluation, and a listing of tests
no longer available. Some of the tests described
are part of the special Head Start Test Collection,
which has been established to provide informa-
tion about instruments for those engaged in
research involving young children. Such tests are
specifically indicated. (NE)

ED 085 531 95 CE 000 709

Delaware's Occupational-Vocational Education
Model: Career Development Data Gathering In-
strumentation.

Delaware State Board for Vocational Education,
Milford.

Spons. Agency—Bureau of Adult, Vocational, and
Technical Education (DHEW/OE), Washing-

ton, D.C.
Bureau No.—0-361-0013
Pub Date Aug 73
Grant—OEG-0-71-0678(361)
Note—69p., For related document, see CE 000 710

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Career Education, *Data Collection, Questionnaires, Resource Materials, Surveys, *Teacher Developed Materials, Test Reviews, *Tests, *Test Selection, Test Validity, Vocational Education

The booklet is an annotated compilation of data gathering instrumentation for use in planning, implementing, and evaluating career education programs. It contains two sections: (1) eighteen instruments developed by Delaware's Occupational-Vocational Education Model project staff for specific use in the local school system and (2) six commercially available instruments. The major criterion for selecting tests, surveys, and questionnaires was how well they measured project objectives. Commercially available instruments were used whenever feasible. An example of each project-developed instrument is included. Addresses of publishers of commercially available instruments are given. Three basic criteria for selecting a data gathering instrument are suggested: they are measurement validity, appropriateness, and administrative usability. (MS)

ED 085 746

CS 200 926

Diederich, Paul B.
What Statewide Testing Can Do.
Pub Date Nov 71

Note—12p., Paper presented at the Annual Meeting of the National Council of Teachers of English (61st, Las Vegas, November 25-27, 1971)

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Curriculum Evaluation, Education, *Educational Assessment, Evaluation, Learning, *State Surveys, *Testing, *Test Interpretation, Test Results

Statewide testing can serve four important functions: can illustrate superior results of a group of schools where no one would expect it and raise questions about how they accomplished it; statewide testing deals with the generally lower scores of disadvantaged minorities, it can put the differences in perspective by showing comparable differences between boys and girls, testing statewide can deal with school effects other than knowledge and basic skills, as shown not only by an interest measure but also by data on attitudes toward school, and it can show that a particular program is producing substantial and socially important results. It is not necessary to give the same test to everybody in the whole state if the objective is to discover the strong and weak points in the state's educational system. (WR)

ED 086 721

TM 003 371

Segel, David
State Testing and Evaluation Programs.
Office of Education (DHEW), Washington, D.C.
Report No.—CIRC-320
Pub Date 51
Note—39p.

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Educational Testing, Program Descriptions, *State Programs, *State Surveys, Student Evaluation, *Testing Programs

Identifiers—*State Testing Programs
This report is concerned with the state testing and evaluation programs which were in effect during the year 1949-1950. Each program account describes the agency which coordinated the program, the nature and purpose of the program including the tests administered, the uses of test results, and other pertinent information, and the publications which were available from each state. A summary of the purposes for the Texas program is provided in the Appendix to illustrate various uses of the results of state-wide testing. States which had no program are so indicated. (NE)

ED 086 730

TM 003 381

Symposium on an Alternative to Traditional Test

Reporting.

Spons Agency—Iowa Testing Programs, Iowa City.

Pub Date 10 Nov 72

Note—55p., Paper presented at Mid-South Research Association, November 10, 1972

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Achievement Tests, *Computer Programs, *Educational Testing, Feedback, Parents, *Reports, Scoring, Standardized Tests, Student Evaluation, Student Testing, Teachers, *Test Results, *Verbal Communication

Identifiers—Madison Project, *Narrative Test Reporting, Test Reporting

This symposium presents an alternative to the numerical mode of reporting, and describes the work that has been done, and the work that is continuing, in the area of computer-generated narrative-format testing reports. Computer narrative testing reports are in a verbal form that eliminates the use of almost all numbers and technical terms. These reports have been designed for the pupil, his parents, and his teacher. The need for alternative reporting systems is discussed in terms of serving the unique needs of the examinee and others who receive reports of his test scores. The application of computer technology to test reporting covers a brief overview of verbal format testing reports and describes the Madison Project with its accomplishments and its limitations. The Iowa Project is discussed in terms of its objectives and the kind of narrative reports that are planned. Each of the audiences of the reports is discussed in terms of their needs and the kinds of reports that are available to them. The presentation, The Future of Narrative Test Reports, sketches the current plans for the Iowa Project and some long-range possibilities and applications for narrative format reports. (END)

ED 086 736

TM 003 391

Cameron, Colin
Discrimination in Testing. Bibliography. Revised, April 1973.

Wisconsin Univ., Madison. Inst. for Research on Poverty.

Pub Date Apr 73

Note—146p.

EDRS Price MF-\$0.65 HC-\$6.58

Descriptors—*Bibliographies, *Disadvantaged Groups, Employment Interviews, *Equal Opportunities (Jobs), *Minority Groups, Predictive Ability (Testing), *Test Bias, Testing Problems

Over one thousand books and articles published between 1942 and 1973 are listed in this bibliography. These citations are concerned not only with discriminatory testing, interviewing and recruiting processes, but also with efforts on the part of employers to overcome these problems. Emphasis is placed on discrimination in the employment and ability testing of adults, rather than academic testing, although important citations of the latter, as well as the testing of children, are included. Many of the entries are annotated. A list of sources for researchers who are interested in further developments on the topic of "Discrimination in Testing" is included. (MP)

ED 087 739

SP 007 720

Testing.

National Education Association, Washington, D.C. Div. of Instruction and Professional Development.

Pub Date 73

Note—44p.

Available from—Instruction and Professional Development, National Education Association, 1201 Sixteenth Street, N.W., Washington, D.C. 20036 (No price quoted)

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Educational Testing, *Guidelines, *Professional Associations, *Standardized Tests, Teacher Certification, Teacher Education, *Testing

Identifiers—*National Education Association, NEA

This is a report of the findings and recommendations of the Division of Instruction and Professional Development of the National Education Association (NEA) on testing. NEA called for a

moratorium on standardized testing in 1972 and created the task force on testing, whose work is summarized in this report. After an introduction stating the problem, the document presents NEA resolutions and new business items on testing. Included in this are statements of task force beliefs, some of which are as follows: a) some measurement and evaluation in education is necessary; b) certain measurement and evaluation tools are either invalid, unreliable, out of date, or unfair and should be withdrawn from use (sharply criticized were standardized achievement and intelligence tests as they affect bilingual/bicultural students; c) the training of those administering tests is inadequate, and schools of education, school systems, and testing industry must take this responsibility; d) there is overkill in the use of standardized tests; and e) the National Teacher Examinations are an improper tool and must not be used for certification, selection, salary determination, tenure, dismissal, and similar matters. The document includes recommendations for immediate action and further study. "The Report of the Committee on Accountability to the NEA Representative Assembly July 1973" and a bibliography. (JA)

ED 087 789

TM 003 397

State Testing Programs: 1973 Revision.

Educational Testing Service, Princeton, N.J.; ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.

Pub Date Nov 73

Note—65p.

Available from—Available from Advisory and Field Services, Educational Testing Service, Princeton, N.J. 08540 for \$4.00

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Educational Assessment, *Educational Programs, *State Surveys, *Testing Programs

Identifiers—State Testing Programs

The purpose of the survey was to obtain information to prepare a profile of state testing programs. One section of the report summarizes the data on the 42 testing programs that were operating in 33 states during the 1972-73 school year. This summary tabulates the findings of eight major areas covering all the questions asked during the interviews, including: purposes of programs; management aspects; population tested; instrumentation; data collection and processing; norms; dissemination; and prospects for the future. Detailed program descriptions for each state are presented in the second section. Finally, the two appendices present an item by item response summary across states and programs and a copy of the interview guide. (MP)

ED 087 791

TM 003 411

Levenson, Hanna

Reliability and Validity of the I.P. and C Scales—A Multidimensional View of Locus of Control.

Pub Date Aug 73

Note—9p., Paper presented at American Psychological Association Convention (Montreal, Canada, August, 1973)

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—Activism, Behavior, *Behavior Rating Scales, Citizen Participation, Dissent, *Locus of Control, Parent Influence, Patients (Persons), *Personality Studies, Personality Theories, Predictive Validity, Psychological Characteristics, Reinforcement, Social Behavior, *Test Reliability, *Test Validity

Identifiers—Rotters Internal External (IE) Scale
The paper contains a discussion of the modification of Rotters' Internal External Scale into the three scales of Internal, Powerful Others, and Chance orientations. The tripartite, multidimensional view of locus of control was developed because the validity and usefulness of Rotters' lumping expectancies of fate, chance, and powerful others together under the rubric of external control were questioned. Data are provided indicating that the refinement of the Internal External scale is justified. The three predicted orientations emerged in factor analyses, and scores on the scales were differentially related to such variables as philosophy of human nature, involvement, information, activism, psychopathology, and perceived parental upbringing. (Author)

Jacobs, Stanley S.

The Evaluation of the Culturally Different: Preschool, Primary and Elementary Age Children.

Pub Date 31 Oct 73

Note—22p.; Paper presented at the 4th Annual Convocation of the Northeastern Educational Research Association, Ellenville, N.Y., October 31-November 2, 1973

EDRS Price MF-\$0.65 HC-\$3.29

Descriptors—*Children, Elementary School Students, Preschool Children, *Standardized Tests, *Student Evaluation, Syntax, *Test Bias, Test Construction, *Testing Problems, Test Reliability, Test Validity

This report asserts that the evaluation of young children can be viewed as a classic case of evaluation of the culturally different. Emphasized is the fact that not only are the majority of tests developed with an adult's perspective concerning adequacy of directions, items, and formats, but also the evaluation of the products is carried out within the context of adult experience and imperfect memory. Selected research literature concerning factors influencing the evaluation of young children is reviewed. Some new data is presented, including a syntactic analysis of verbal directions for children's tests and an analysis of the difficulty of directions read by children on other instruments. The normative equivalents of change-level scores on a number of standardized measures of cognitive variables are examined. The evidence supports the original conceptualization of cultural bias. (Author/NE)

Journal Articles

EJ 065 887 040 CG 505 010
The Minnesota Vocational Interest Inventory
 Clark, Kenneth E. *Measurement and Evaluation in Guidance*, v5 n3, pp418-419, Oct 72
 *Vocational Interests, *Test Interpretation, *Measurement Instruments, *Vocational Counseling, *Counselor Role, Counselors
 The Minnesota Vocational Interest Inventory provides scores that directly relate to many of the issues counselors and counselees discuss. The counselor who has such information at his disposal can give meaningful assistance to an individual facing a decision regarding his future career. (Author)

EJ 066 049 090 CG 505 012
The Career Planning Program—More than a Test
 Battery Hanson, Gary R.; Cole, Nancy S. *Measurement and Evaluation in Guidance*, v5 n3, pp415-419, Oct 72
 *Career Planning, *Tests, *Vocational Counseling, *Counselor Role, Counseling, Testing, Testing Programs
 In response to Goldman's allegation in volume 4 of this journal that testing and counseling are not complementary, the author describes the Career Planning Program which represents one attempt to improve the relationship between tests and counseling. (Author)

EJ 066 387 180 CG 505 009
The Basis for a Lacking Relationship Between Tests and Counseling
 Layton, Wilbur L. *Measurement and Evaluation in Guidance*, v5 n3, pp403-407, Oct 72
 *Tests, *Counseling, *Testing, *Counselor Role, *Test Interpretation
 Goldman's conception of the relationship between tests and counseling based on his remarks in an article in volume 4 of this journal is rejected because of the surplus meaning in the marriage analogy. (Author)

EJ 066 388 180 CG 505 013
It's Time to Put Up or Shut Up
 Goldman, Leo. *Measurement and Evaluation in Guidance*, v5 n3, pp420-423, Oct 72
 *Tests, *Counseling, *Counselor Role, *Measurement, *Measurement Goals, Testing
 Tests and measurement have fallen short of the contributions to counseling expected of them. The writer takes the position that inadequacies of both the tests and counselors are responsible and that major changes are needed if tests are to continue to have a place in the counseling process. (Author)

EJ 066 395 180 CG 505 087
The Effect of Verbal Approval Upon the Performance of Middle- and Lower-Class Third-Grade Children on the WISC
 Galdieri, Anthony A.; And Others. *Psychology in the Schools*, v9 n4, pp404-408, Oct 72
 *Socioeconomic Status, *Lower Class Students, *Performance Factors, *Intelligence Tests, *Testing, Elementary School Students, Testing Problems
 Results of the study indicate that regardless of the presence or absence of verbal approval by the examiner during the administration, there was no significant difference in children's test performance on the WISC. (Author)

EJ 067 404 520 CG 505 011
OVIS-A Nonpredicting Device
 D'Costa, Ayres G. *Measurement and Evaluation in Guidance*, v5 n3, pp411-414, Oct 72
 *Tests, *Vocational Counseling, *Testing, *Measurement Instruments, Prediction, High School Students
 The Ohio Vocational Interest Survey was designed to serve primarily as an occupational exploration device for high school students. The OVIS profile presents measured interests alongside vocational plans and thus facilitates an individual learning process with which the counselor can assist. OVIS emphasizes a marriage between testing and counseling. (Author)

EJ 067 405 520 CG *05 082
A Word of Caution on the Use of the WAIS
 Walker, Kenneth P.; Walker, Carol A. *Psychology in the Schools*, v9 n4, pp374-378, Oct 72
 *Intelligence Tests, *Tests, *Standardized Tests, *Retardation, Testing Problems

EJ 068 094 190 CG 505 128
Comparison of Satisfied and Dissatisfied Users of Holland's Self-Directed Search
 Collins, Anne M.; Sedlacek, William E. *Journal of Counseling Psychology*, v19 n5, Sep 72
 *Occupational Guidance, *Self Evaluation, *Vocational Counseling, *Tests, *Counseling Effectiveness, College Freshmen
 This study provides some evidence that individuals may find the Self Directed Search differentially satisfactory in vocational counseling. (Author)

EJ 068 112 190 TM 500 771
The Relationship of Each of Sex Scales of the Study Attitudes and Methods Survey (SAMS) to Each of Two Criteria of Academic Achievement in a Community College
 Miller, Doris Crane; Michael, William B. *Educational and Psychological Measurement*, v32 n4, pp1107-110, W 72
 *Academic Achievement, *Community Colleges, *Predictive Validity, *Measurement Instruments, Factor Structure, Predictor Variables, Multiple Regression Analysis, Tables (Data), [*Study Attitudes and Methods Survey, SAMS]

EJ 068 113 190 TM 500 776
An Instrument to Measure Vocational Maturity
 Westbrook, Bert W.; And Others. *Educational and Psychological Measurement*, v32 n4, pp1131-3, W 72
 *Vocational Adjustment, *Measurement Instruments, *Test Validity, *Test Reliability, Test Construction, [*Vocational Maturity, Cognitive Vocational Maturity Test, CVMT]
 This report describes the Cognitive Vocational Maturity Test (CVMT), and instrument designed to measure career knowledge and abilities within six areas of the cognitive domain of vocational maturity, as well as offers validity and reliability data. (Authors)

EJ 069 064 520 TM 500 766
Relationships between the California Test of Mental Maturity and the Stanford Achievement Test Battery at the Primary Levels
 Merenda,

Peter F.; And Others. *Educational and Psychological Measurement*, v32 n4, pp1079-87, W 72
 *Aptitude Tests, *Achievement Tests, *Comparative Analysis, *Measurement Instruments, Primary Grades, Correlation, Tables (Data), [*Canonical Correlation Analysis, California Test of Mental Maturity, Stanford Achievement Test]

EJ 069 066 520 TM 500 770
Myers Achievement Motivation Scale: A Validation Study
 Stewin, L.; Nyberg, V. *Educational and Psychological Measurement*, v32 n4, pp1103-6, W 72
 *Test Validity, *Motivation, *Rating Scales, *Measurement Techniques, Secondary School Students, Intelligence Quotient, Grade Point Average, Tables (Data), [*Myers Achievement Motivation Scale, Canada]
 Several differences existed between the sample described in Myers' study (Educational and Psychological Measurement, 1965, 25, 355-363) and that of the present study, particularly concerning educational aspirations and environmental factors. (Authors/CB)

EJ 069 076 010 CG 505 180
On Black Intelligence
 Williams, Robert L. *Journal of Non-White Concerns in Personnel and Guidance*, v1 n1, pp9-14, Oct 72
 *Intelligence, *Negroes, *Achievement Tests, *Cultural Factors, *Test Bias, Intelligence Tests, Culture Free Tests
 The author concludes that the continued administration of traditional ability tests to Black children without correcting for the cultural bias is a violation of the child's constitutional rights. (Author)

EJ 069 708 180 CG 505 246
Test Review
 Gable, Robert K. *Journal of Counseling Psychology*, v19 n6, pp565-570, Nov 72
 *Tests, *Work Attitudes, *Test Reviews, Evaluation, [*Work Values Inventory]
 The reviewer concludes that the Work Values Inventory is not psychometrically adequate for gathering vocational information relevant to an individual's vocational decision. Donald Super, author of the WVI, comments on the review.

EJ 070 289 340 AA 514 598
Factor Analysis of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI)
 Hollenbeck, George P.; Kaufman, Alan S. *Journal of Clinical Psychology*, v29 n1, pp41-5, Jan 73
 *Factor Analysis, *Factor Structure, *Intelligence Tests, *Test Validity, Early Childhood, Tables (Data), [*Wechsler Preschool & Primary Scale of Intelligence, WPPSI]
 Results were quite similar for each of the factor analytic procedures, an indication of a strong underlying structure of the WPPSI tests. Regardless of the factor analytic technique used, two factors appeared—one clearly verbal and the other clearly performance—at each of the three age levels. (Authors)

EJ 070 815 040 CG 505 387

Sexual Bias in Interest Measurement Harmon, Lenore W. *Measurement and Evaluation in Guidance*, v5 n4, pp496-501, Jan 73

*Interest Tests, *Sex Differences, *Bias, *Occupational Choice, *Test Bias, Testing, Test Construction, Measurement

Bias in interest testing occurs if tests are used to encourage an individual to consider, enter, or reject an occupation or type of occupation on the basis of an irrelevant variable, sex. In general, the externally referenced inventories, which reflect the current state of affairs and have well established predictive powers, have the most potential for sexual bias. (Author)

EJ 071 004 090 CG 505 419
Some Useful Tests for Marriage Counseling Phillips, Clinton E., *Family Coordinator*, v22 n1, pp43-53, Jan 73

*Marriage Counseling, *Tests, *Evaluation, *Adult Counseling, *Personality Assessment, Family Counseling, Interpersonal Relationship, Marital Instability

Many who do marriage counseling are unaware of the vast amount of test materials which are available to them or of their tests which might well aid them and their clients in the process of counseling. Some widely used tests in marriage, family, and premarital counseling are discussed, and resources are listed. (Author)

EJ 071 391 180 CG 505 351
Measurement in Counseling Research: A Review Zytowski, Donald G., Betz, Ellen L., *Counseling Psychologist*, v3 pt. 2 n4, pp72-79, 72

*Research, *Measurement, *Counseling, *Measurement Instruments, *Research Reviews (Publications), Counselor Characteristics, Psychometrics, Counseling Effectiveness, Rating Scales

A review of several leading counseling journals and some books reporting major research projects was performed, covering material from 1961 to the present. Instruments were sorted as follows: (1) Counselor characteristics measurements, and (2) Counseling relationship and effectiveness measurements. Also examined were areas of reliability and validity research. (CJ)

EJ 072 202 450 AA 514 695
Immediate Knowledge of Results and Test Performance Beeson, Richard O., *Journal of Educational Research*, v66 n5, pp224-26, Jan 73

*Academic Achievement, *Student Evaluation, *Test Results, *Test Interpretation, *Analysis Of Variance, Educational Research, Statistical Analysis, Reinforcement

The purpose of this study was to investigate the possible effect of students having an immediate awareness of the results of tests they have taken and how that knowledge influences further test performance. (Author/RK)

EJ 073 955 520 AA 515 048
Tests Have Many Uses But Beware the Abuses Instructor, v82 n7, pp50-2, Mar 73

*Standardized Tests, *Test Results, *Student Testing, *Test Interpretation, *Student Evaluation, Manuals, Scores, Test selection, Achievement Tests

Analyzes the uses that can be made from test data and the influences such data can have on communication with communities, in the evaluation of school's effectiveness, and student's placement. (Author/RK)

EJ 077 171 520 AA 514 264
For Millions of Children, the United Teaching Profession Says "No" to Statewide Testing Lehrhaupt, Arthur, *NJEA Review*, v46 n4, pp24-5, 27, Dec 72

*Testing, *Group Intelligence Testing, *Testing Programs, *State Government, State Programs, State Legislation, Teachers, [New Jersey] believes that the New Jersey Commission of Education's Policy of state-wide testing creates an "atmosphere of repression against teachers" and "curbs imagination and innovation" in children (DS)

FJ 078 793 520 RC 501 223
Measurement of Intelligence and Language Differences Vasquez, James, *Aztlan*, v3 n1, pp155-63, Spr 72

*Intelligence Tests, *Language Tests, *Mexican Americans, *Nonverbal Tests, *Test Bias, Culture Free Tests, Educational Philosophy Translated and nonverbal tests are discussed as techniques for measuring intelligence and language differences in the Chicano child. (NQ)

EJ 079 508 190 TM 500 908
Relationships Between the EPI Scales and the 16 PF CPI, and EPPS Scales Edwards, Allen L., Abbott, Robert D., *Educational and Psychological Measurement*, v33 n2, pp231-8, Sum 73

*Measurement Instruments, *Rating Scales, *Personality Assessment, *Factor Structure, *Comparative Analysis, Tables (Data), [Edwards Personal Preference Schedule, California Psychological Inventory, 16 Personality Factor Questionnaire, Edwards Personality Inventory]

Study was undertaken to determine the degree to which the EPPS, CPI, and 16 PF scales are correlated with the EPI scales and to determine the degree to which the scales in all four inventories are measuring the same common traits. (Authors)

EJ 081 158 180 CG 505 992
The Wechsler in Personality Assessment: Object Assembly Subtest as Predictive of Bodily Concerns Stewart, Denton J., and Others, *Journal of Consulting and Clinical Psychology*, v40 n3, pp488, June 73

*Research, *Personality Assessment, *Psychological Tests, *Psychological Evaluation, *Test Interpretation, Psychological Studies

Research is contradictory concerning ability of the Wechsler Object Assembly (OA) subtest to predict bodily preoccupations. This study presents data from an objective personality test that do not support the hypothesis that bodily concerns are indicated by impaired performance on the WAIS OA subtest. (Author)

EJ 081 177 180 UD 502 498
A Note on Minority Group Test Bias Studies Reilly, Richard R., *Psychological Bulletin*, v80 n2, pp130-132, Aug 73

*Test Bias, *Minority Groups, *Ethnic Studies, *Sociocultural Patterns, *Statistical Analysis, Culturally Disadvantaged, Self Concept, Cross Cultural Studies

EJ 081 915 420 PS 502 729
A Locus of Control Measure for Preschool Children Stephens, Mark W., Delys, Pamela, *Developmental Psychology*, v9 n1, pp55-65, Jul 73

*Psychological Studies, *Preschool Children, *Measurement Techniques, *Question Answer Interviews, *Expectation, Correlation, Early Experience, Parent Influence, Data Analysis, Reinforcement

This paper (a) describes the rationale and development of the Stephens-Delys Reinforcement Contingency Interview method for assessing

locus of control expectancies in children as young as preschool age and (b) summarizes results of subsequent studies bearing on reliability, validity, and problems of the method. (Author/RK)

EJ 082 088 490 CG 505 881
Revision and Validation of the Trux-Carkhuff Relationship Questionnaire Lin, Tien-Teh, *Measurement and Evaluation in Guidance*, v6 n2, pp82-86, July 73

*Relationship, *Interpersonal Relationship, *Counseling Effectiveness, *Counselor Role, *Measurement Techniques, Counselor Evaluation

A demand for a short, valid, and reliable scale to measure the counseling relationship was met by revising and validating the Trux-Carkhuff Relationship Questionnaire. (Author)

EJ 082 289 AA 516 385
An Evaluation of Forced-Choice and True-False Item Formats in Personality Assessment Jackson, Douglas N., and Others, *Journal of Research in Personality*, v7 n1, pp21-30, Jun 73

*Personality Assessment, *Evaluation, *Forced Choice Technique, *Psychological Studies, *Peer Groups, Females, Questionnaires, Data Analysis, Tables (Data), Correlation

The reliabilities and validities of true-false and forced-choice formats in personality assessment were compared. (Author)

EJ 082 454 CG 506 023
Relationships Among the Edwards Personality Inventory Scales, the Edwards Personality Preference Schedule, and the Personality Research Form Scales Edwards, Allen L., Abbott, Robert D., *Journal of Consulting and Clinical Psychology*, v40 n1, pp27-32, Feb 73

*Research, *Personality Tests, *Test Results, *Test Interpretation, *Correlation, Evaluation The present study was undertaken to determine the degree to which the EPPS and PRF scales are correlated with the EPI scales, and also to determine the degree to which the scales in all three inventories are measuring the same common traits. (Author)

EJ 082 616 EA 504 031
The IQ Puzzle: What Are We Measuring? Kagan, Jerome, *Inequality in Education* n14, pp5-13, Jul 73

*Testing Problems, *Test Bias, *Intelligence Quotient, *Intelligence Differences, *Minority Groups, Public Schools Discusses weaknesses and cultural biases of intelligence tests. (JF)

EJ 082 617 EA 504 032
Separate and Unequal Again Green, Winifred, *Inequality in Education* n14, pp14-16, Jul 73

*Testing Problems, *Ability Grouping, *Equal Education, *Minority Groups, *Racial Discrimination, School Integration, Public Schools

Discusses how testing and tracking affect students. Argues that grouping reinforces years of discriminatory treatment in the education of black children, locking them into classroom situations in which the stigmas and hopelessness are the same or worse than in the days of separate but unequal schools. (JF)

EJ 083 237 SE 508 984
Problems in the Use of Educational Tests Jackson, Paul H., *Mathematical Spectrum*, v4 n2, pp49-59, 71/72

*Educational Testing, *Mathematical Models, *Test Reliability, *Test Validity, Classification, Mathematics, Measurement, Prediction, Sampling, True Scores
A brief survey of the mathematics involved in the theory of educational testing. (MM)

EJ 083 600 UD 502 517
IQ Tests and Their Educational Supporters
Jorgensen, Carl C., *Journal of Social Issues*, v29 n1, pp33-40, W 73

*Intelligence Tests, *Test Bias, *Negro Students, *Test Validity, *Administrator Responsibility, Testing Problems, Educational Diagnosis, Test Reliability, Educational Accountability, Test Interpretation

IQ tests are used to judge the mental capacities of black Americans, a fact which places a tremendous ethical responsibility on those involved in IQ test research and administration; most IQ test developers and administrators have abrogated this responsibility. (Author/JM)

EJ 084 104 CG 506 145
A Comparison of Wechsler Preschool and Primary Scale of Intelligence and Stanford Binet Intelligence Scale Scores for Disadvantaged Preschool Children Anthony, John J., *Psychology in the Schools*, v10 n3, pp297-299, Jul 73

*Preschool Children, *Disadvantaged Youth, *Intelligence Level, *Intelligence Tests, *Measurement Techniques, Intelligence, [Wechsler Preschool and Primary Scale of Intelligence, Stanford Binet Intelligence Scale]

This comparative study of these two instruments indicates that they yield similar IQ scores when used in the assessment of disadvantaged preschool children. (Author)

EJ 084 106 CG 506 147
Identification of Diagnostic Strengths and Weaknesses on the McCarthy Scales of Children's Abilities Ysseldyke, James E.; Samuel, Simon, *Psychology in the Schools*, v10 n3, pp304-307, Jul 73

*Diagnostic Tests, *Measurement Techniques, *Children, *Ability Identification, *Identification, Ability

EJ 084 182 CG 506 223
Development Of An Effective Screening Protocol For Identifying Learning Disability Students Jackson, Frank, *Texas Personnel and Guidance Association Journal*, v2 n2, pp137-139, Sep 73

*Prediction, *Identification Tests, *Learning Disabilities, *Screening Tests, *Measurement Instruments, Elementary School Students, Urban Schools, Suburban Schools
Study results suggest that a practicable combination of screening instruments can identify those children with a learning disability. The syllabication test and the teacher rating scale can be used with an 80 percent effectiveness level. (Author/CJ)

EJ 084 248 CG 506 289
The Effects Of Audio-Visual Test Presentation Berry, Gene A., And Others, *College Student Journal*, v7 n2, pp76-78, Apr-May 73

*College Students, *Testing, *Student Testing, *Measurement Techniques, *Test Construction, Testing Programs
The purpose of the present experiment was to test whether a combination of audio and transparency projection, transparency projection, or the standard paper and pencil method of testing is superior when testing college students on a multiple-choice achievement test in basic psychology. (Author)

EJ 085 222 SO 502 086
Ability Tests? A Shot in the Dark Petrovsky, Arthur V., *Prospects*, v3 n2, pp179-181, Sum 73

*Educational Theories, *Psychological Testing, *Educational Testing, *Test Interpretation, *Predictive Ability (Testing), Measurement Techniques, Testing Problems, Test Reliability, Predictive Validity
Several areas of controversy between Soviet psychologists and their Western colleagues concerning the usefulness of tests for measurement of mental ability are noted in this article. The author outlines test procedures suggested by Russian psychologist, Lev Vygotsky. (SM)

EJ 085 721 CG 506 366
The Measurement of Cognitive Vocational Maturity Westbrook, Ben W.; Parry-Hill, Joseph W., Jr., *Journal of Vocational Behavior*, v3 n3, pp239-251, Jul 73

*Employment Qualifications, *Occupational Information, *Vocational Maturity, *Work Environment, *Measurement Instruments, Test Reliability, Cognitive Development, Vocational Development, Test Construction, [Cognitive Vocational Maturity Test]

Describes an instrument designed to measure an individual's level of cognitive vocational maturity in six areas: Fields of Work, Job Selection, Work Conditions, Education Required, Attributes Required, and Duties. When vocational choices were in agreement with field of interest and ability level, scores on the subtests were higher. Mean scores on all subtests increase across grade levels, supporting the claim that vocational maturity behaviors are developmental ones. (Author/EK)

EJ 085 727 CG 506 372
A Scale to Measure Attitudes Toward Working Alfano, Anthony M., *Journal of Vocational Behavior*, v3 n3, pp329-333, Jul 73

*Employment, *Unemployed, *Work Attitudes, *Measurement Instruments, *Rating Scales, Semiskilled Workers, Skilled Workers, Unskilled Workers
A scale, developed to measure an individual's attitude toward work and the relationship between length of unemployment time and work attitudes, was administered to subjects who were: (1) employed workers—in unskilled, semi-skilled, and occasionally skilled jobs; (2) unemployed but actually seeking employment; (3) chronically unemployed. The scale differentiated these groups readily, establishing that a relationship exists between work attitudes and length of unemployment. (Author)

EJ 085 961 EA 504 227
Women and Educational Testing Tittle, Carol Kehr, *Phi Delta Kappan*, v55 n2, pp118-119, Oct 73

*Sex Discrimination, *Test Bias, *Educational Testing, *Feminism, *Test Construction, Testing Problems, Equal Education
Presents documentation suggesting sex bias can be found in educational testing and suggests that a more objective treatment of women in educational tests can be made by showing women in a wider variety of occupations and activities and by more equal representation of women in test content. (Author/DN)

EJ 086 070 EC 060 012
Data Gathering: The Reliability and Validity of Test Data from Culturally Different Children Adler, Sol, *Journal of Learning Disabilities*, v6 n7, pp429-34, Aug-Sep 73

*Exceptional Child Education, *Culturally Disadvantaged, *Testing Problems, *Test Reliability, *Test Validity, Disadvantaged Youth, Examiners

EJ 089 264 CG 506 592
A Comparison Of Urban And Rural Reliability Estimates For The Boehm Basic Concept Test Houch, Cherry K.; And Others, *Psychology in the Schools*, v10 n4, pp430-432, 73

*Concept Formation, *Rural Urban Differences, *Test Reliability, *Socioeconomic Influences, *Cognitive Tests, Primary Grades, [Boehm Basic Concept Test]
This study determined reliability of the Boehm Basic Concept Test (designed to assist the primary teacher assess basic conceptual knowledge) for rural kindergarten and first grade children. They found the rural reliability to be less than that reported for urban samples especially for middle and upper class children. The reduced reliability was accompanied by reduced variance in the rural middle and upper class groups. (DK)

EJ 089 638 HE 504 843
Analysis of an Innovation (CLEP) Caldwell, Edward, *Journal of Higher Education*, v44 n9, pp698-702, Dec 73

*Higher Education, *Testing, *Tests, *Test Reliability, *Equivalency Tests, Educational Certificates, Evaluation Techniques, Student Evaluation

Since credit-by-examination is an intrinsically sound educational device, those who administer it might well take steps to protect it from abuse. Colleges should insist that both content specialists and measurement experts participate in evaluating the tests and the College Entrance Examination Board should revise its procedures so that institutions can easily get information to evaluate the tests. (Author)

EJ 089 459 EA 504 476
Legal Challenges to Educational Testing Practices Weckstein, Paul, *Inequality in Education* n15, pp92-101, Nov 73

*Court Litigation, *Ability Grouping, *Racial Discrimination, *Equal Protection, *Tests, Test Validity, Test Bias, Elementary Schools, Secondary Schools

Schools frequently use test scores to divide students into groups, i.e., to classify them. In attempting to challenge such procedures, lawyers can draw on a developed body of case law on testing in education, employment, and other areas. Discusses these cases. (JF)

EJ 089 714 HE 504 919
A Reminder for Educators: Seven Points to Remember About the SAT Angoff, William H., *College Board Review* n90, pp5, W 73-74

*Higher Education, *Testing, *Measurement Techniques, *Student Testing, *Tests, Achievement Tests, Measurement Instruments, [Standard Achievement Tests]

EJ 089 806 IR 500 033
Individual Testing As A Guide Brethower, Karen S., *Improving Human Performance*, v2 n3, pp205-210, F 73

*Individual Instruction, *Material Development, *Measurement Techniques, *Test Validity, *Instructional Design, Instructional Technology, Instructional Materials, Evaluation Criteria, Testing, Programed Instruction

Testing instructional materials with individual subjects produces unique and invaluable data for revising the instructional materials. Examples of the procedures and the usefulness of individual testing demonstrate the critical importance of the technique in developing validated instruction. (Author)

EJ 090 066 SE 510 038
A Flexible and Forgiving Testing Program. A Report of the Current Status of the Testing Program of the CRIMEL Project Elbrink, Larry C. *School Science and Mathematics*, v73 n8, pp686-690, Nov 73

*College Mathematics, *Mathematics Education, *Program Descriptions, *Testing Programs, Educational Programs, Individualized Instruction, Mathematics, Testing, [CRIMEL Project]

Outlines the structure of the CRIMEL project in mathematics at the Ohio State University. Describes the testing program which gives students the opportunity to retake examinations if they believe that they can improve on their initial performance. (JR)

EJ 090 357 TM 501 056
Personality Assessment at the Individual Level Using the Semantic Differential Everett, A. V. *Educational and Psychological Measurement*, v33 n4, pp837-44, W 73

*Measurement Techniques, *Semantic Differential, *Personality Tests, Individual Characteristics, Self Concept Tests

The feasibility of personality measurement through the Semantic Differential and an examination of the degree of relationship between the semantic space and the preference and similarity space derived by independent means are reported. All three phases of the study met with essentially negative results. (Author/NE)

EJ 090 360 TM 501 059
Internal Consistency of the Coopersmith Self-Esteem Inventory Spatz, K. Christopher, Johnston, James O. *Educational and Psychological Measurement*, v33 n4, pp875-76, W 73

*Self Esteem, *Self Concept Tests, *Test Reliability, Rural Schools, Racial Attitudes, Grade 5, Grade 9, Grade 12, [Coopersmith Self Esteem Inventory]

EJ 090 377 TM 501 076
The Reliability and Predictive Validity of a Test of Independence of Work Habits Parc, Roland R., Butzow, John W. *Educational and Psychological Measurement*, v33 n4, pp963-65, W 73

*Predictive Validity, *Test Reliability, *Independent Study, *Predictive Measurement, Academic Achievement, Success Factors, Study Habits

EJ 090 380 TM 501 079
Validity Concerns of Preschool Testing Pascale, Pietro J. *Educational and Psychological Measurement*, v33 n4, pp977-78, W 73

*Preschool Children, *Test Validity, *Measurement Techniques, *Preschool Tests, Comparative Testing, Test Reliability

EJ 090 420 UD 502 752
Race of Subject and Race of Examiner Jacobs, John F., DeGraaf, Carl A. *Integrated Education*, v11 n4, pp68-71, Jul-Oct 73

*Race Influences, *Intelligence Tests, *Testing Problems, *Examiners, *Test Interpretation, Student Teacher Relationship, Psychological Testing, Testing Problems, Educational Diagnosis, [Wechsler Intelligence Scale for Children]

Reports a study which investigated the influence of the variables of race (examiner and child race) and expectancy (high and low expectancy) upon the scoring of individual IQ tests. (Author/JM)

EJ 090 853 CG 506 626
Comparison of Three Methods of Reducing Test Anxiety: Systematic Desensitization, Implosive Therapy, and Study Counseling Cornish, Richard D., Dilley, Josiah S. *Journal of Counseling Psychology*, v20 n6, pp499-503, Nov 73

*Anxiety, *Desensitization, *Study Skills, *Tests, *Academic Performance, Study Habits, College Students, Grade Point Average, Grades (Scholastic)

Systematic desensitization, implosive therapy, and study counseling have all been effective in reducing test anxiety. In addition, systematic desensitization has been compared to study counseling for effectiveness. This study compares all three methods and suggests that systematic desensitization is more effective than the others, and that implosive therapy is more effective than study counseling. (Author)

EJ 090 926 CG 506 709
A Simple Vocational Interest Inventory For Counseling Meir, Elchanan I. *Vocational Guidance Quarterly*, v22 n2, pp126-128, Dec 73

*Interest Tests, *Vocational Counseling, *Occupational Aspiration, *Career Choice, *Vocational Interests, Aspiration, Interest Research, Occupational Tests

This article shows how a client's level of aspiration and vocational interests can be measured by means of a short, reliable, and valid interest inventory. (Author)

EJ 090 956 CG 506 739
Effects Of Knowledge Of Ability Test Results On Academic Performance And Test Anxiety McMahon, Michael P. *Journal of Counseling Psychology*, v20 n3, pp247-249, May 73

*Test Results, *Academic Performance, *Anxiety, *Feedback, *Test Interpretation, Grade Point Average, College Students, Standardized Tests, [School and College Ability Test, Test Anxiety Scale]

The analysis of grade point averages showed no significant differences among detailed, limited, or no knowledge or results groups. The test anxiety level was higher for the detailed knowledge group. It was concluded that the time factor between giving the results to subjects and the grade point average collection and administration of the test anxiety measure could have reduced the impact of the knowledge on the results. (Author/LA)

EJ 091 033 CS 700 529
An Investigation of Potentially Harmful Effects of "Neurological Organization" Procedures on Children with Brain Dysfunction Krippner, Stanley. *Reading World*, v13 n2, pp92-112, Dec 73

*Neurologically Handicapped, *Psychological Studies, *Child Language, *Social Behavior, *Behavior Change, Mentally Handicapped, Language Development, Parent Child Relationship

The assertion that Neurological Organization treatment procedures are potentially harmful is not supported by the data in this study, although negative results of such treatment on the child's linguistic and social behavior and on the functioning of the family unit has been frequently conjectured. (Author/RB)

Doctoral Dissertations

FACULTY ATTITUDES TOWARD STANDARDIZED TESTING IN THE PUBLIC SCHOOLS OF PENNSYLVANIA

Robert Barmont CORMANY, D.Ed.
The Pennsylvania State University, 1973

The purpose of this study was to examine the attitudes of teachers and counselors in the public schools of Pennsylvania toward the use of standardized tests. These attitudes were compared on the basis of the quality of the standardized testing programs, the size of the student populations and the financial expenditures of the participating school districts and a number of variables descriptive of the respondents such as sex, position and grade level.

A sample of 45 school districts was selected and from the staff of these districts 1206 teachers and counselors were chosen by random sampling to participate in the study. Each subject completed the Faculty Attitudes toward Standardized Testing Inventory, an attitude measure developed by the author for this study, consisting of 45 items which the subjects responded to on a four-point, Likert scale. A total of 1010 completed inventories were returned. At the same time a quality index was computed for each district's testing program on the basis of information obtained from a guidance program survey form completed by each district for the Pennsylvania Department of Education in 1971.

The data were treated by an analysis of variance technique using The Pennsylvania State University's computer center. Specific hypotheses shown to exhibit significant variance among the mean attitude scores for their cells were then investigated further by means of a Scheffé test.

It was found that the financial expenditures of a school district did have a direct relation to the attitudes of teachers and counselors toward the standardized testing program, particularly for school districts with more than 5000 students. However, the size of the total student population within a district did not appear to influence these attitudes significantly. In the case of the quality index that had been computed for each district's testing program, no significant relationship was found between it and the attitudes of the teachers and counselors as measured by the attitude inventory. The respondents' attitudes toward standardized testing were found to vary significantly at the $p < .01$ level for the demographic variables of sex, position, grade level, highest degree earned, previous service on a test advisory committee and opinions as to the degree of information possessed concerning the district standardized testing program. No significance was detected for the variables of number of years in education and number of courses taken in the area of tests and measurement.

Order No. 73-24,000. 140 pages.

A TEST OF THE VALIDITY OF HOLLAND'S PERSONALITY THEORY OF VOCATIONAL CHOICE

Susan Dunn DeVOGE, Ph.D.
University of North Carolina at Chapel Hill, 1973

Supervisor: Dr. Luther R. Taff

The purpose of this study was to further investigate the validity of Holland's personality theory of vocational choice. The first two stages of the study were an attempt to duplicate Holland's research and to investigate its validity by using The Sixteen Personality Factor Questionnaire (Cattell and Eber, 1962). The first step was to investigate whether major area

groups of college students (based on Holland's classification scheme) could be significantly predicted on the basis of the students' personality scores as freshmen. The second step was to investigate whether the major area groups of students could be predicted based on the students' personality scores as seniors. It was felt that this would make it possible to better assess the validity of Holland's theory and of the importance of personality variables in major area choice. In stage three, a further investigation of Holland's theory was undertaken by conducting a follow-up study on these same subjects four years after their graduation. By analyzing a questionnaire which obtained occupational information, it was possible to determine the proportion of subjects who were employed in an occupation whose environmental type matched their major area type. It was felt that determining whether continuity existed between major field of study and occupational choice was an important addition in assessing Holland's theory.

The population used in this study consisted of 132 male subjects who were students at the University of North Carolina at Chapel Hill from 1965 to 1969. The group was randomly selected in the fall of 1965. The Sixteen Personality Factor Questionnaire was administered to them as freshmen and again in 1969 as seniors. The students' major areas were obtained at the end of their senior year. These were classified into Holland's six environmental types using his classification scheme for major area fields. These personality data were analyzed by using discriminant analysis to determine whether Holland's major area groupings could be predicted on the basis of the students' personality scores as freshmen and as seniors. Occupational follow-up data were obtained by questionnaire for these same subjects four years after graduation. The subject's occupation was classified by using Holland's classification scheme for occupations. Each subject's major area type was compared with his occupational type to determine the proportion of subjects who selected occupations in the same environmental area as was their major field of study.

Analysis of the data showed that it was not possible to significantly differentiate between the major area groups of students based on their personality scores as freshmen. It was possible, however, to significantly discriminate between the major area groups of students based on their personality scores as seniors. A percentage table was constructed and it was found that the number of subjects whose occupational type matched their major area type was significantly different from chance. Two of the three stages of the study, therefore, provided support for Holland's personality theory of vocational choice. It was concluded that the overall study provided moderate support for Holland's theory.

Order No. 74-5912, 93 pages.

THE DEVELOPMENT AND PRELIMINARY TESTING OF THE VOCATIONAL LEVEL SURVEY

Randall Weston RICHARDS III, Ed.D.
Rutgers University The State University of New Jersey, 1972

The objective of the investigation was to test on a preliminary basis an instrument designed to measure the level of vocational development of college-bound adolescents. The Vocational Level Survey represented an attempt by the investigator to operationalize certain attitude and behavior continua developed by Donald E. Super and others in order to measure vocational development.

A sample of 300 typical college-bound students was selected to test the instrument during the period from December, 1971,

through early February of 1972. The students were drawn from grades nine through twelve and college freshmen. Completed surveys from 237 students were evaluated. The participants were 105 boys and 132 girls. The ages of the subjects ranged from 13.5 years to 24.4 years with a mean at 16.4 years. Biographical data, career planning information, and academic aptitude test score results were also collected on the subjects.

The hypotheses for this investigation were generated as a result of a review of expected correlates of vocational maturity which have been studied by other researchers (Crites, 1969; Dilley, 1965; Montesano & Geist, 1964; Osipow, 1968; Super et al., 1957, 1960; Westbrook, 1971). These hypotheses dealt with age, grade level, extracurricular activities, achievement and socioeconomic status of the subjects.

A computer generated intercorrelation matrix provided Pearson Product-Moment values for all of the variables in the study. An examination of the matrix indicated a range of low ($r = .17-.24$) but significant relationships ($p < .01$) between VLS scores and the first three hypotheses (age, grade level, and extracurricular activities). Some limited support was found for the hypothesized relationships between the remaining two areas (scholastic aptitude and socioeconomic status) and total scores on the VLS.

The Spearman-Brown coefficient of reliability ($r = .72$) for the VLS which resulted from this preliminary investigation was encouraging. Hence, additional testing and evaluation of the instrument may be warranted.

An item analysis of the VLS statements indicated that about 70 per cent of the items did discriminate between the high and low scores on the Survey. In addition the importance of factors related to planning and commitment in vocational development were reaffirmed as indicated by a range of low ($r = .35-.37$) yet significant ($p < .01$) relationship between certain items on the personal data sheet and total VLS scores.

The results of the data analysis suggested that further refinement of the instrument was needed before it could be used as a counseling tool. Nevertheless, the findings of the study tended to agree with Super's view of the process of vocational development.

In retrospect, the VLS appeared to be measuring characteristics of the vocational development process within the limits of the investigation. In addition the relative importance of information getting and planning activities in the adolescent's repertoire of vocationally relevant thoughts and behaviors was indicated.

The implications of the study seemed to be that certain attitudes and behaviors contributed significantly to the vocational development of the college-bound adolescents who participated in the investigation. It may be suggested that further research could be directed at determining which information getting and planning activities are best able to serve the needs of maturing adolescents. The identification of specific groups of activities could provide counselors with meaningful ingredients for use in counseling adolescents with problems of vocational development.

Order No. 73-6441, 109 pages.

A STUDY OF REPORTED RELATIONSHIPS BETWEEN INTELLIGENCE AND SCHOOL ACHIEVEMENT AS MEASURED BY STANDARDIZED TESTS

Joseph Anton TWARANOVICA, Ed.D.
University of Illinois at Urbana-Champaign, 1973

The purpose of this study was to summarize and, in so far as possible, synthesize the reported relationships between the results of intelligence and school achievement as measured by standardized psychometric devices.

The data came from a total of 410 studies. Through a succession of orderings, the data for each type of educational designation were generated into a series of tabulations of the correlations for facets of tested intelligence and achievement. With regard to information concerning sex differences, similar tabulations were fashioned. The correlation data were then analyzed

in terms of the median, range, and number of correlations for each of the series of values tabulated. These sets of correlational information, in turn, were arranged in terms of three basic dimensions: aspects of achievement, kinds of intelligence devices (group, individual) and levels of educational organization (Single, Combined, General Educational).

Unfortunately poorly described studies dominated the literature. Many of the authors of these studies were at a mass stage in their perceptions of the testing of intelligence and school achievement. All too frequently they focused on the kinds of tests in a global undifferentiated manner. Some didn't even deign to provide the name of the intelligence test(s) they used. Many neglected to indicate the kind of measure -- IQ or MA -- which they used in obtaining their correlations. A parallel laxity in description was demonstrated with respect to the tests of educational achievement. There frequently was failure to differentiate among such achievement facets as reading comprehension, vocabulary, arithmetic reasoning, computation, and the like. Little apparent awareness was in evidence of a sensitivity to a difference among the measures associated with achievement. The bulk of the reported correlations were made with a mass type of measure -- the nonspecified "score" -- used for the subject-matter device. The populations were all too often inadequately delineated, terms such as "elementary level" (with no grade level specified) being used, and surprisingly few authors showed any sensitivity to the desirability of examining their phenomena in terms of sex differences.

In spite of the distressing conditions of the reported research sampled, only eight percent regarded as yielding well differentiated findings, certain observations seemed warranted. (1) The reported correlations between measured intelligence (group and individual tests) and components of school achievement spanned a wide range. Yet these values showed a discernible tendency toward being positive. (2) Generally, a perceptible increase over single grade segments of the school was noted for the magnitude of the median correlations between measured group learning aptitude and subject-matter areas. (3) A discernible upward tendency in the magnitude of the median correlations, between group intelligence tests and subject-matter areas for combined grade levels was evident. (4) Clearcut directional trends in correlational magnitude were not apparent for general educational grade ranges, measured group intelligence and achievement components. (5) Information relevant to the magnitude of the relationships over the various grade designations when measured individual intelligence and facets of achievement were considered, were too limited for directional determination. (6) Some slight differences in degree of relationships occur in different subject-matter areas when group learning aptitude tests and single grade levels were examined. These differences were not readily understandable in terms of the kinds of behavior samplings made for various facets of achievement. (7) Differing results were often probably a function of the nature of the population utilized; variations among devices; and attenuating conditions such as the kind of measures used. (8) A paucity of data, within restricted grade ranges, and widely fluctuating values reported for boys and girls prevented meaningful comparison with regard to sex differences: across grade levels or between grade levels. Order No. 74-5721, 328 pages.

PRACTICES OF LARGE PUBLIC SCHOOL SYSTEMS CONCERNING THE PUBLIC RELEASE OF STANDARDIZED ACHIEVEMENT TEST RESULTS

Lyle Osborn YOUNG, Ed.D.
The University of Tulsa, 1973

Director: Professor Marlow A. Markert

This study investigated the practices of large public school systems concerning the public release of standardized achievement test results: which systems released test results for the 1970-1971 school year and what specific practices were utilized in the public dissemination. The study sought to deter-

mine the reasons, advantages, and disadvantages of the dissemination policies, and to investigate the future of testing and public dissemination as anticipated by superintendents.

All public school systems in the nation with pupil memberships of 50,000 or more were surveyed, with responses received from seventy of the eighty systems. Returns indicated that 46 per cent of the responding systems did not release 1970-1971 standardized achievement test results to the public. Twenty-six per cent released only system-wide results, while 28 per cent released school-by-school results in addition to the system-wide results. The survey revealed that there is disagreement among systems concerning the dissemination policies, and that the policies are in a state of flux.

The policies were examined by sorting the systems by size, per cent non-white membership, and geographic location to determine if such variables influenced the dissemination practices. The greatest difference in dissemination policy between systems was by geographic region, but the factor associated with these regional differences was state mandated testing programs. Almost half of the systems that released test results to the public were in California, Colorado, and Florida which have state testing programs.

Generally, accompanying data were released with the achievement test results. Almost three-fourths of the schools that released results to the public reported some type of related data such as racial composition, class size, per pupil expenditure, dropout rate, or follow-up data about graduates.

Generally, systems utilized a combination of reporting methods. Achievement test results were usually released in a special report at a board of education meeting with the news media receiving copies or summaries of the data presented. Reading test results were most often released, with mathematics and language also frequently released. Results were most often released at the elementary level. Both grade equivalent scores and percentile ranks were utilized at the elementary and junior high levels, with the percentile rank most frequently used in releasing results for the senior high level. The respondents indicated that the test publishers' national standardization norms were always used in releasing the results to the public.

The reason most often stated for withholding test results was to avoid misinterpretation of the results by the press and the public. The reason most frequently reported for releasing results was that the board of education, public, or legislature had demanded the disclosure of results. The advantage most often stated for withholding results was to avoid erroneous conclusions about high and low scores. However, the disadvantage most often stated was that the public had lost confidence in the schools. The advantage generally given for releasing results was that public interest and involvement in the schools had increased. The disadvantage most often stated was that disclosure leads to misinterpretation and speculation by the public and press regarding low or high scores.

When asked to speculate about the future of testing, the opinion most frequently presented was that testing would continue, and public disclosure of school-by-school test results would increase. According to the respondents the accountability movement, with its demands for hard data to prove that children are learning, will provide the impetus for the increase in public disclosure.

Order No. 74-2835, 117 pages.

A COMPARATIVE ANALYSIS OF THE PERFORMANCE OF CLINIC SUBJECTS ON THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN AND THE REVISED ILLINOIS TEST OF PSYCHOLINGUISTIC ABILITIES

Robert Michael STAUSING, Ed.D.
Ball State University, 1973

Adviser: Dr. Robert E. Hayes

The purpose of this study was to investigate the amount and patterns of interrelatedness of the Wechsler Intelligence Scale

for Children (WISC) and the 1968 revised edition of the Illinois Test of Psycholinguistic Abilities (ITPA). The study attempted to analyze from both the practical and statistical levels relationships between the two instruments. Further clarification was sought as to the: (1) dimensions measured by the ITPA, (2) concurrent validity of the ITPA, and (3) necessity of administering both instruments in their entirety in psychological evaluations.

Subjects for the study were 236 children who had been referred to three clinics located in central Indiana. The basis of selection for the study was that the child had been: (1) administered at least the WISC and ITPA within a six month period, and (2) referred for evaluation of non-specified learning difficulties. There were 179 male and 57 female subjects included in the study. They ranged in age from four years and nine months to ten years and three months.

Data for each subject subsequently submitted for statistical treatment included sex, test age, scaled scores for each sub-scale administered and the WISC Full Scale IQ. Data not subjected to statistical treatment were those in which the numbers were insufficient for the application of statistical procedures. Data eliminated from final statistical treatment were race, socio-economic status, predicted Stanford-Binet IQ and Psycholinguistic Quotient.

Statistical treatment to ascertain the relationship between the WISC and the ITPA included the calculation of: (1) correlation coefficients by the product-moment method from ungrouped data, and (2) canonical correlations. A variance ratio (F) was used for testing the significance of the computed product-moment correlations. Chi-square values were used for testing the statistical significance of each canonical correlation. The degree of significance necessary to reject the null hypothesis that the two tests were uncorrelated was set at the .05 level. In order to attain practical significance a correlation of $\pm .40$ was required.

Comparing the mean and standard deviation of each WISC-ITPA subtest for the study population to the original normative samples for the instruments it was found that: (1) the WISC subscales show slightly less variability, and (2) the ITPA subscales show more variability. Product-moment correlations between the variables ranged from $-.255$ for Vocabulary on the WISC and sex, to $.670$ for the WISC Full Scale IQ and the Auditory Association subtest of the ITPA. Two significant canonical correlations were extracted from the data. The first appeared to be composed of auditory memory and oral language elements. The second, a bipolar correlation, was composed of positive visual components and negative auditory-vocal elements.

Analysis of the data revealed that significant statistical and substantial practical degrees of relationship exist between scaled scores of the WISC and the ITPA. The null hypothesis was rejected since the relationship existing between the WISC and the ITPA variables exceeded the critical point at the .05 level for the canonical correlations and at or beyond the .05 level for the product-moment correlations.

Based on comparative analysis of the WISC and ITPA the following conclusions appear warranted:

1. There exists a highly significant degree of statistical, and a substantial degree of practical relationship between the instruments.
2. The pattern of highly significant correlations found to exist between the instruments add to the concurrent and predictive validity of the individual ITPA subtests, and the test as a whole.
3. When the subject responds at or near age levels on the WISC the Auditory Association, Grammatical Closure and Auditory Sequential Memory subtests may be deleted from the ITPA administration.

4. The "single abilities" concept of the ITPA authors appears untenable, with the exception of the Sound Blending subtest.

Order No. 74-1730, 87 pages.

TEACHER KNOWLEDGE OF STANDARDIZED TEST INFORMATION AND ITS EFFECT ON PUPIL I.Q. AND

ACHIEVEMENT

Feige SOROTZKIN, Ph.D.

Case Western Reserve University, 1973

The primary purpose of the present study was to examine the effect upon pupil performance of providing second grade teachers with various types of standardized test information within the context of the teachers' views toward testing. The intellectual and academic performance of the second grade students were defined as the dependent variables and were assessed by means of pupil performance on the Kuhlman Anderson Test and the Stanford Achievement Test. The independent variables involved in the present investigation were test information condition, teacher opinion of tests and sex of student.

The sample consisted of 567 second grade boys and girls from predominantly black lower class inner city schools. Twenty three teachers and six schools were involved in the study. Measures of standardized intelligence and achievement were obtained at the beginning of the school year. Students were then randomly assigned to one of four test information treatment groups: (1) Stanford Achievement Test scores made available to teacher, (2) Kuhlman Anderson I.Q. scores made available to teacher, (3) Stanford Achievement and Kuhlman Anderson I.Q. scores made available to teacher, and (4) no standardized test information made available to teacher. Class lists providing the teacher with varying kinds of standardized test information for each member of the class were distributed. In May of the same school year, the standardized achievement and intelligence measures were readministered.

The teacher opinion construct employed in this study was based upon the work of Fleming and Anttonen (1971) and refers to the attitudes and opinions of teachers about tests and the abilities they allegedly measure. Designation of teacher opinion group was assigned on the basis of teacher response to the four selected items of the Teacher Opinion Index. Teachers were categorized as having either high-middle or low opinions of standardized tests.

The results of an analysis of covariance model showed no significant differences in pupil I.Q. and achievement as a function of the kind of test information distributed to the teacher. The analysis further revealed no significant differences in pupil I.Q. between students whose teachers had high-middle opinions of tests and those whose teachers had low opinions of tests.

Significant findings resulting from the statistical analyses for pupil I.Q. and achievement included: (1) second grade students whose teachers had high-middle opinions of tests scored significantly higher ($p < .05$) on final Vocabulary Achievement and on Arithmetic Achievement ($p < .01$) than second grade students whose teachers had low opinions of standardized tests, (2) the final I.Q. scores of second grade girls differed significantly ($p < .05$) from those of the second grade boys, (3) second grade girls scored significantly higher ($p < .01$) on final Word Reading Achievement, Spelling Achievement and Word Study Achievement than second grade boys. Significant differences at the .05 level were also noted for Paragraph Meaning Achievement. The final Vocabulary Achievement scores of the boys significantly exceeded those of the girls ($p < .01$) and presented an interesting departure from the well documented research in the field.

The results of the present study suggest that test information does not appear to have a significant effect on the intellectual functioning and academic achievement of second grade students. This finding raises serious questions about the usefulness of test information in the academic world. The significant results that were evident as a function of teacher opinion and sex of student support the idea of a broad expectation network operating in the classroom which takes into account the multi-faceted characteristics of teachers and students.

Order No. 74-2567, 144 pages.

Stephen St. Clair TAYLOR, Ph.D.
The Florida State University, 1973

Major Professor: Irwin R. Jahns

The purpose of this study was to evaluate the effects of training adult basic education teachers in behavioral attending skills as test administrators. The need for such a study evolved from the underlying problem of adult test anxiety. Specifically, the primary objective was to determine if an increase in the use of behavioral attending skills by adult basic education teachers would lead to a corresponding decrease in test anxiety. An additional objective was to examine the relative permanence of the learned behavioral skills acquired by the teachers in the experimental group of this investigation.

The research population consisted of fifteen randomly selected adult basic education teachers from the Central Adult Campus of Florida Junior College in Jacksonville, Florida.

A one-to-one pre-training testing session was video taped while each teacher administered a four minute timed test to a randomly assigned adult examinee. After the pre-training testing session the eight teachers in the experimental group participated in a microtraining session that lasted one hour and consisted of the following three phases: (1) Didactic. Each teacher was asked to read an attending behaviors manual that described the basic behaviors of attentiveness. During this phase the teachers were encouraged to discuss any facet of the information contained in the manual. (2) Modeling. "Good" and "bad" examples of the use of the individual behaviors of attentiveness were modeled by the trainer. Each teacher then practiced the behaviors that were modeled. (3) Experiential. The teachers viewed the video tape recordings of their own pre-training testing session with his peers. The trainer reinforced each teacher with approval for both appropriate use of the behaviors as well as for accurate identification of the instances.

A subsequent one-to-one post-training testing session was then video taped with all fifteen subjects. This was followed, two months after the microtraining session, with a third testing session.

In sum, the fifteen adult basic education teachers were video taped three times each in a one-to-one testing situation, with different adult examinees each time, across a two-month time span. After each testing session the adult examinee was asked to complete a rating scale designed to quantify testing environment anxiety. These ratings, along with the video tapes that were analyzed for attending behaviors by a team of raters, were the dependent measures. These measures were then statistically analyzed by using a one-way analysis of variance on the pre-training data and an analysis of co-variance, with test one as the covariate, for the post-training and follow-up data.

With alpha set at .05, the following null hypotheses were rejected:

1. The frequency of test examiner attending behaviors, as rated by a team of raters, will not differ significantly between the experimental and control groups after training the experimental group of adult basic education teachers in the behavioral skills of attending.

2. Test anxiety, as perceived and reported by the adult examinee, will not differ significantly between the experimental and control groups after training the experimental group of adult basic education teachers in the behavioral skills of attending.

3. The frequency of test examiner attending behaviors, as rated by a team of raters, will not differ significantly between the experimental and control groups at the time of the follow-up study two months after the training session.

4. Test anxiety, as perceived and reported by the adult examinee, will not differ significantly between the experimental and control groups at the time of the follow-up study two months after the training session.

Order No. 74-8413, 90 pages.

A STUDY OF THE EFFECTS OF MICROTRAINING FOR ATTENDING BEHAVIORS IN ADULT TESTING

THE PERFORMANCE OF LOWER CLASS BLACK AND LOWER CLASS WHITE CHILDREN ON THE WEPMAN AUDITORY DISCRIMINATION TEST: THE EFFECTS OF DIALECT AND TRAINING, AND THE RELATIONSHIP TO READING ACHIEVEMENT

Gertrude Wernick KARGER, Ed.D.
Harvard University, 1973

The present study examined three questions: 1) What is the effect of Black English on the auditory discrimination performance of lower class black children? 2) What is the effect of training on the auditory discrimination performance of lower class black and lower class white children? 3) What is the relationship of auditory discrimination performance to reading achievement?

The Wepman Auditory Discrimination Test and Supplementary Auditory Discrimination Test were administered to 60 low SES black and 60 low SES white first grade children, randomly selected and assigned to one of the two treatment groups of trained/untrained. The untrained children received the standardized instructions presented in the Wepman manual. The trained children were given guidance in the use of the terms "same" and "different" as well as practice with final consonant discriminations. Reading achievement was measured by means of the Stanford Achievement Test (Word Reading and Paragraph Meaning) and by means of two oral reading tests constructed by the author - Reading Tests A and B. The Slosson Intelligence Test and Goodenough-Harris Draw a Man Test were also administered.

An analysis of covariance for the 30 "different" Wepman items, with IQ as covariate, indicated that the black children scored significantly lower than the white children, irrespective of training. At the outset of the study, four possible sources of bias against the Black English speaker in the 30 "different" items were identified: a) the final [θ - l/-v] pairs, b) the final stop pairs, c) the individual item pen-pin, and d) the "miscellaneous" final consonant pairs. Analyses of covariance indicated that 1) the black children scored significantly lower than the white children on (a), (b), and (c); 2) that the difference approached significance on (d); 3) that there were no significant differences between the black children and white children on any other groups of items. These results were interpreted as an indication that Black English is in part responsible for the poor performance of low SES black children on the Wepman test.

With respect to training it was found that on the 30 "different" Wepman items the trained children scored significantly higher than the untrained children. The adjusted mean of the white trained group was adequate while that of the white untrained group was inadequate. Consequently, it was concluded that low SES white first grade children are not "deficient" in auditory discrimination ability. The adjusted mean of the black trained group, although higher than that of the black untrained group, still did not reach the adequate level - a further indication of the effect of Black English. The training effect on the Wepman test was greater for the final than the initial consonant pairs. In addition, for the 30 "different" pairs and for the 13 final consonant pairs, training helped the black children and white children to an equal extent. However, for the six final stop pairs, training helped the black children more than the white children.

The analyses of covariance for the Supplementary test indicated that on the "dialect" pairs (homonyms in Black English) and "miscellaneous" final consonant pairs, the black children scored significantly lower than the white children; on the "contrast" pairs (heard as contrasting words in Black English) there was no difference in performance between the black children and white children. These results indicated that the difficulties which Black English speakers experience in discriminating word pairs are predictable and related to the phonological system of Black English.

The following relationships were found between auditory discrimination performance and reading achievement for the white children: For the untrained children, Wepman performance predicted performance on the reading tests primarily

evaluating word recognition (Stanford Word Reading and Reading Test A), even with the effect of IQ controlled. The results for the trained group were not as clear since the correlation between the Wepman test and Reading Test A - with IQ held constant - was significant while that between the Wepman test and Stanford Word Reading was not. The Stanford Paragraph Meaning subtest was found to be related to the Wepman test to a lesser extent than the word recognition tests. With the effect of IQ controlled, neither the correlation for the trained nor the untrained group was significant.

For the black children - both trained and untrained - Wepman performance predicted performance on the reading tests which primarily evaluated word recognition (Stanford Word Reading and Reading Tests A and B), even with the effect of IQ controlled. On the other hand, with IQ held constant, the correlation between the Wepman test and Stanford Paragraph Meaning was not significant for either the trained or untrained group. These results, along with the conclusion arrived at earlier that Black English affects the Wepman performance of low SES black children, were interpreted as an indication that Black English has an influence on the acquisition of word recognition skills by low SES black children.

Order No. 74-11,323, 361 pages.

THE DEVELOPMENT OF A SPANISH LANGUAGE VERSION OF THE SLOSSON INTELLIGENCE TEST

POOLE, Sister M. Julianna, Ed.D.
The University of Rochester, 1974

The research which has been described in this dissertation was undertaken in order to explore and experience some of the problems which are inherent in the testing of Spanish-speaking children. The dearth of adequate psychometric instrumentation in Spanish has been decried by educators for many years. For this reason, the author elected to develop an individually administered test of intelligence in Spanish. Because of its brevity, ease of administration and scoring, and proven reliability and validity, the Slosson Intelligence Tests (SIT) was chosen as a base from which this instrument was developed.

Sixty first-grade subjects with Spanish surnames who attended public schools of the City School District of Rochester, New York, and Catholic schools within the district, were tested. A simple random sample was drawn from each of these two sub-populations with relative sample size proportional to relative sub-population size. This proportion was approximately 3 (Catholic) to 10 (public).

Items in the SIT are equated with chronological age levels. To insure that an adequate middle range of items for first-grade subjects would be available, questions from the third year (3-0) to the fourteenth year (14-0) were translated. As a result of a feasibility study conducted prior to the actual test administration, the test floor was lowered to 2-0. In all, 90 items were translated.

Subjects were pre-screened by means of a measure which incorporated 15 short commands in Spanish. The administration of the pre-screening measure and the translated portion of the SIT - Escala de Inteligencia Slosson (EIS) - took about 30 minutes. All testing was done by the author who is bilingual.

The range of the raw scores obtained by the 60 subjects was 24 to 53 with a median score of 35, a mean of 36.35, and a standard deviation of 5.97. Twenty-eight of the subjects were boys. Their mean score was 33 with a mean of 34.85 and a standard deviation of 6.64. The median score for the 32 girls was 37 with a mean of 37.64 and a standard deviation of 4.97.

An analysis of the items answered by all 60 subjects indicated that 65% of the items proved to be relatively easy and an equal percentage did not discriminate well, if at all, between high and low groups. These results could have been due to the low test floor and consequent ease of the preliminary items.

Total test and item reliability were found by means of the test-retest method. Thirty of the subjects were retested within a two-day interval. A correlation of .93 was obtained between

scores on both testings, with a standard error of measurement of 1.47. When the scores of boys and girls were analyzed separately, the correlations were .95 and .90, respectively. Item reliability was measured by the phi coefficient. The typical item reliability was not very satisfactory. The generally low correlations might have been the result of differential effects introduced by the short interval between test and retest.

The validity of the EIS was examined by correlating raw scores of subjects on this test with those obtained on achievement tests in Mathematics and Reading in English. The correlations for total test and subtest scores were very low. An analysis of possible reasons for these low correlations suggested that further research was necessary before conclusions as to the validity of the *Escala de Inteligencia Slosson* could be made.

An analysis of test results showed that sex differences in favor of the girls were evident. Recommendations were proposed concerning the reranking of items, changes to be made in several items, the administration of items at different age levels, the use of longer test-retest intervals, and further comparisons between the EIS and tests in Spanish and English.

The research reported in this study is descriptive of an initial phase in test development. The difficulties accompanying such an endeavor are apparent. It is the intention of this investigator to continue searching for explanations and solutions to the problems associated with adapting a test from one language to another.

Order No. 74-20,629, 128 pages.

THE BENDER GESTALT TEST AMONG MEXICAN CHILDREN

RIVERA, Felipe U., Ph.D.

The Catholic University of America, 1974

Very little research has been done with the Bender Visual Motor Gestalt Test among children from different cultures. The objectives of this investigation are: (1) to find out the "normal" performance of Mexican children on the Bender Test, (2) to establish developmental norms for that population, and (3) to compare the performance of Mexican and American children on the same test.

The subjects for this study are 1284 urban children proportionally selected by age, sex, type of school (private or public), and by geographic location of the schools. The Bender Test was individually administered according to the Koppitz (1964) testing instructions, and her developmental system was used to score all the protocols.

Mean scores were established for the groups of this sample, and t-tests were used to determine the significance of the differences between the means. Similar differences were determined between Mexican and American children on the Bender Test performance. Total item analysis was done for all 1284 protocols of the Mexican sample at every 12 month interval, and the percentage of error was reported for each item. These proportions were compared with those proportions of American children (Snyder & Snyder, 1974), and Z scores were used to establish the significance between them.

It was found that boys perform better than girls, and that private school children perform better than public school subjects. It was found too, that American children perform at a higher level than Mexican children. These differences indicate that the Bender Test performance is modified by cultural factors, and that the Koppitz Developmental Norms are not appropriate for Mexican children.

Further research is needed to validate this test against other criteria, such as intelligence tests, school performance and brain dysfunction within the Mexican population.

Order No. 74-19,426, 166 pages.

CONCURRENT VALIDITY OF THE AMERICAN COLLEGE TESTING INTEREST INVENTORY FOR MINORITY GROUP MEMBERS

LAMB, Richard Ray, Ph.D.

The University of Iowa, 1974

Supervisor: Professor Albert B. Hood

This study provides information on whether vocational interest inventories are appropriate for minority group members. Criterion-related (concurrent) validity of the American College Testing Interest Inventory (ACT IV) was compared between national samples of white college seniors and college seniors in four minority ethnic groups -- Black, American Indian, Spanish American, and Oriental American. Samples used were described by responses to questionnaire items requesting information on field of study, goals, satisfaction with major, changes in major, grades, and source of funds. Comparisons were made on the ability of the ACT IV to correctly classify white seniors as to field of study with its ability to correctly classify seniors in each of the minority groups. Additional tests for cultural bias in inventory items were made through comparison of white and minority samples on scale reliabilities (internal consistency) and intercorrelations of the six Holland-type scales used in the ACT IV. Agreement between distributions of actual and predicted fields of study was compared between samples to determine whether minority individuals had been consistently "channeled" into some fields inconsistent with their interests. Results indicated concurrent validity was not substantially lower for any of the minority samples for either sex. Scale reliabilities were equivalent for all samples. The scale intercorrelation matrix for only one sample -- American Indian males -- was found to differ substantially from that of the white subjects. This finding suggested a possibility of cultural bias in the ACT IV for American Indian males. Distributions of actual and predicted fields of study were, in general, similar for all samples. Implications of results for other inventories and younger minority populations were discussed. Order No. 75-1216, 144 pages.

THE EFFECT OF SIMULATED TESTING SESSIONS ON WORKSAMPLE PERFORMANCE SCORES OF BLIND INDIVIDUALS

FINKLE, Louis Joseph, Ph.D.

George Peabody College for Teachers, 1974

Major Professor: Randall K. Harley

The purpose of the study was to investigate the effects of a series of simulated testing sessions on subsequent worksample performance scores of blind individuals. The subjects consisted of 17 blind students ranging in age from 15 to 23 years. The sample was randomly divided into an experimental group (n = 9) and a control group (n = 8). The experimental group participated in the simulated testing sessions, while the control group maintained their daily school schedule. It was hypothesized that there would be significant differences between the postintervention worksample scores of the experimental and control groups. The differences would favor the experimental group due to treatment effects.

The intervention consisted of a series of 1-hour per day simulated testing sessions conducted for a period of 9 days. The experiment was structured as a randomized posttest-only design. The data was statistically analyzed using t tests, analysis of variance, and analysis of covariance. Postintervention

worksamples used to measure the basic industrial processes of assembling, sorting, packaging, and producing, included the Pennsylvania Bi-manual Test, a sorting criterion, The Clawson Envelope Worksample Test, and the Minnesota Rate of Manipulation Turning Test respectively.

The results disclosed significant improvement of scores of the experimental group on the Pennsylvania Bi-manual Test. The intervention did not appear to have a significant effect on the experimental group's scores using the Sorting criterion, Clawson Envelope Worksample Test, and the Minnesota Rate of Manipulation Turning Test. It was concluded that blind individuals perform significantly better on worksamples involving the industrial process of assembling if exposed to a series of simulated testing sessions. The simulated testing sessions had no statistically significant effect on the subjects' performance of the industrial processes of sorting, packaging, and producing. The most limiting covariate was grade level. Due to possible effects of grade level on the results, the sorting and packaging scores were nonsignificant between groups.

The results have implications for testing, training, and serving the general population. By referring to the general population who may lack worksample experiences, one may see the need for further experimentation in simulation sessions using worksample theory. The primary advantage of active learning may be realized through more extensive use of worksamples in educational programming.

Order No. 74-29,166, 90 pages.

THE CONSTRUCTION AND VALIDATION OF LANGUAGE-RELATED TESTS OF SPEECH-SOUND DISCRIMINATION FOR SPECIFIC LANGUAGE POPULATIONS IN FIRST AND SECOND GRADES

BRYEN, Diane Nelson, Ph.D.
Temple University, 1973

The Problem

In recent years special education has been seriously criticized as being a dumping ground for culturally different children. Because placement into special classes occurs only after evaluation by a school psychologist, the standardized evaluation instruments used have also been particularly criticized on the basis of their being standardized on white middle-class populations. For many black and Puerto Rican children the use of these instruments threatens their performance because these instruments assume that all potential users share a common culture with white middle-class youngsters and are speakers of standard English. This assumption further threatens their educational careers, for when culturally different children perform poorly on these instruments, the general explanation is that they are linguistically and/or cognitively deficient and many are subsequently placed into special classes.

The purpose of this research was to develop three tests of the sound discrimination of phonological structures in standard and black English, and Spanish. Hypotheses were concerned with the relative performance of white, black and Puerto Rican children on each of the three language forms, when scores on all the forms were combined and on another measure of speech-sound discrimination, Wepman's Auditory Discrimination Test (ADT). Other hypotheses were concerned with age and sex differentiation in each language group on the performance on the appropriate language form.

Procedures

Item analysis of an initial pool of 65 items for black and standard English and 50 items for the Spanish form resulted in revised parallel language forms, with 35 items in each form. The revised forms were administered to 32 blacks, 32 white and 32 Puerto Rican children, with 16 children (eight males and eight females) of each group in both first and sec-

ond grades. These three groups were statistically equivalent with respect to chronological age and socioeconomic status. Each language form was presented to all groups in a counter-balanced order on audio tape. In addition, all subjects were administered the ADT and the Slosson Intelligence Test (SIT). Thirty-one white and 30 black and Puerto Rican children were retested after one to two weeks with the standard English, black English and Spanish forms, respectively.

Results

Significant results included the findings that each language group performed better on the language form that most closely approximated their specific phonological structures, and that when scores on all three language forms were combined there were no differences in speech-sound discrimination ability. As predicted, there were no sex differences and the expected age differentiation was supported for only the Spanish form. Although each language form had a higher correlation with the ADT than the SIT, the differences between these correlation coefficients were not significant. Significant and moderate test-retest reliability coefficients were obtained for each language form.

Discussion

The results of this research support both the linguistic difference model as an explanation of language behavior and the construct validity of these tests of speech-sound discrimination. Possible explanations for the undemonstrated age differentiation on the standard and black English forms centered on differences in intelligence between first and second graders and the narrow age range sampled. Lack of complete support for establishing the relationship between each form and the ADT was explained because of the small to moderate differences in the correlation between each language form and the ADT, and each form and the SIT. Implications for education and recommendations for further research were discussed.

Order No. 74-28,338, 191 pages.

CLASSICAL TEST THEORY FOR CRITERION-REFERENCED TESTS

LIVINGSTON, Samuel A., Ph.D.
The Johns Hopkins University, 1974

Several conflicting definitions of "criterion-referenced" (CR) testing exist. Their authors disagree as to whether CR testing implies random sampling of items, a minimum passing level, or content homogeneity of the test. Nearly all agree, however, that a CR test, unlike a "norm-referenced" (NR) test, is intended to compare examinees not with each other but with an absolute standard. The author defines a CR test as any test for which a "criterion score" is specified independently of the scores for which it will serve as a reference point. A minimum passing level is thus a special case of a criterion score.

The concepts of variance, covariance, and correlation can be generalized to apply meaningfully to CR test score distributions, by redefining them in terms of deviations from the criterion score instead of from the mean. This redefinition preserves the properties of variance and covariance under addition and scalar multiplication of variables, and preserves the bounds of the correlation coefficient.

Classical test-score theory can then be developed in a more general form that allows any point on the scale of measurement to be the reference point. CR reliability can be defined in a manner analogous to NR reliability and expressed in terms of NR reliability. The Spearman-Brown prophecy formula and the correction for attenuation hold for CR reliability in their familiar form; coefficient alpha takes on a modified form. CR reliability can also be defined in terms of variance components.

The estimate of CR reliability from variance components is equivalent to the estimate derived from inter-item covariances.

CR reliability increases with the difference between the mean and the criterion score, because this difference represents reliable information about the examinees' performance in relation to the criterion score. When the mean equals the criterion score, NR and CR reliability are equal; otherwise CR reliability is greater. NR reliability can thus be considered a special case of CR reliability. CR tests can be validated by correlational methods, provided all variables are CR and CR correlation is used. Order No. 74-29,013, 80 pages.

THE EFFECT OF TEST ANXIETY ON THE RELATIONSHIP BETWEEN INTELLIGENCE AND READING ACHIEVEMENT IN A GROUP OF SEVENTH GRADE SUBURBAN CHILDREN

LOMBARDI, Vincent A., Ed.D.
Temple University, 1974

Many researchers have reported negative relationships between test anxiety and reading achievement. In addition, negative correlations have been found between test anxiety and intelligence. These cognitive abilities, while negatively related to test anxiety, are positively related to each other. This long established positive relationship has not been examined at various levels of test anxiety.

The negative relationship between reading achievement and anxiety has previously been examined at various levels of intelligence. There have been mixed results reported. In this study the relationship was reexamined at various levels of intelligence.

The 152 male and 150 female students in this investigation comprised the entire seventh grade population of a suburban junior high school near Philadelphia. They were administered the appropriate forms of the Science Research Associates, Primary Mental Abilities Test, the Metropolitan Achievement Test (Advanced Reading) and the Test Anxiety Scale for Children.

Pearson product-moment correlations were computed between reading achievement and intelligence test scores at high, moderate and low test anxiety. The correlations were compared by the standard z-transformation. Pearson product-moment correlations were computed between test anxiety and reading achievement at low, moderate and high intelligence. The correlations were compared by the standard z-transformation.

The linear correlations between reading scores and intelligence for the total population was .73. The relationships between reading scores and intelligence were .80 for the high anxious group, .67 for the moderate anxious group, and .68 for the low anxious group. The comparisons indicated that the correlation for the high test anxious group was significantly higher (.05) than those of the low or moderate test anxious groups.

The correlation between reading achievement scores and test anxiety scores for the population was -.30. When the negative correlations between reading achievement scores and test anxiety scores were compared at high, moderate and low levels of intelligence the groups did not differ significantly.

It was concluded that the high levels of test anxiety were a significant influence on the relationship between reading achievement and intelligence. The high test anxious student's behavior suggests a motivation to do well on cognitive measures but his concern seems to contribute to lower reading achievement. It was further concluded that the negative relationship between test anxiety and reading achievement was not influenced by low, moderate and high levels of intelligence. Test anxiety would seem to be debilitating to all students regardless of mental ability. Order No. 74-28,181, 86 pages.

AN INVESTIGATION OF THE PREDICTIVE VALIDITY OF THE PEABODY PICTURE VOCABULARY TEST WHEN ADMINISTERED TO PRESCHOOL CHILDREN ENROLLED IN

COMPENSATORY EDUCATION PROGRAMS

RENNER, Anne Healy, Ed.D.
The University of Tennessee, 1974

Major Professor: Dr. Paul C. Burns

The purpose of this study was to investigate the relative predictive value of the Peabody Picture Vocabulary Test (PPVT) when used as an evaluative instrument in a compensatory preschool program for culturally disadvantaged children. One hundred and nine children enrolled in such a program in Buffalo, New York, were administered the PPVT at the beginning of their preschool experience. At the conclusion of the sample's fifth year of formal schooling (fourth grade), the scores from five standardized tests administered as part of the public school's regular testing program were obtained for 82 children from the original sample.

A product-moment coefficient of correlation was obtained for the predictor variable, the PPVT, and the five criterion variables which were the: (1) Lee-Clark Readiness Test; (2) Metropolitan Readiness Test; (3) New York State Reading Test; (4) New York State Arithmetic Test; and (5) Otis-Lennon Mental Ability Scale.

Analysis of the data revealed that the scores obtained on the PPVT had a relatively high positive correlation with the scores obtained on the five other standardized tests. The highest correlation was found between the PPVT and the Otis-Lennon Mental Ability Test which was the final measure of academic achievement obtained for this sample. All coefficients of correlation for the dependent variable and the independent variables were significant at the .0001 level.

Although it was concluded that the PPVT is probably a valid instrument to use in predicting the academic success of disadvantaged children, the weaknesses and dangers inherent in such a practice were discussed. The investigator also presented some specific recommendations to insure the judicious use of standardized tests with culturally disadvantaged children.

Order No. 74-27,233, 173 pages.

A STUDY OF THE RELATIONSHIP BETWEEN NORM-REFERENCED TESTS AND CRITERION-REFERENCED TESTS

VAN VALKENBURGH, Marilyn W., Ed.D.
Western Michigan University, 1974

This study was designed to investigate the relationship between criterion-referenced tests and norm-referenced tests. The questions to be answered were:

1. What is the relationship between norm-referenced tests and criterion-referenced tests with respect to predicting student performance scores or grade equivalents as indicated on a norm-referenced test?
2. What is the relationship between scores on norm-referenced tests and criterion-referenced tests for (a) fourth- and seventh-grade students, (b) black students and white students, and (c) male and female students?

3. What information can the criterion-referenced test provide educational decision makers in decisions pertaining to placement, diagnosis, assessment, prediction, and evaluation?

Both the Metropolitan Achievement Test (norm-referenced test) and the 1973-74 Michigan Educational Assessment Program (criterion-referenced test) were administered to all fourth- and seventh-grade students in a public school system in the Fall of 1973. The sample consisted of 969 fourth-grade students and 949 seventh-grade students. Only those students who were present for the administration of both tests were included in the sample.

The results of the tests were recorded, and several correlation coefficients were computed, including the Pearson and Index of Correlation. All correlations reported were relatively high and consistent, with most of them ranging from .70 to .81 across

tests and subgroups.

It was concluded that the relationship between the criterion-referenced test and the norm-referenced test was of sufficient strength that approximate grade equivalent scores can be predicted from a criterion-referenced instrument. Consequently, it is possible for a criterion-referenced test to provide information to educational decision makers for decisions pertaining to placement, diagnosis, assessment, prediction, and evaluation.

Order No. 74-28,853, 98 pages.

DEVELOPMENT OF CRITERION-REFERENCED TESTS: AN EMPIRICAL INVESTIGATION

STONE, Vathsala Iyengar, Ph.D.
The Florida State University, 1974

Major Professor: Jacob Gordon Beard

This study was concerned with the functional meaningfulness of parallel forms of a Criterion-Referenced (CR) test which were developed according to a pre-postulated 'item sampling procedure'. A performance universe was defined according to the Briggs' model for instructional design and a corresponding item universe was defined by writing an item form for each behavioral objective. Two parallel forms of the test were produced by generating items randomly from each item form and assigning them randomly to each test form. The resulting test forms were evaluated in a Generalizability (G) study using a sample of 320 persons. G coefficients applicable to randomly-parallel as well as stratified-parallel test families were computed for the test forms. The former coefficients were found to underestimate the latter and confirmed the heterogeneity of test content. The range of values for the coefficients were of moderate size. The corresponding variance of the error Δ (i.e., the discrepancy between the observed and the universe scores) was found to be fairly high. The tests were not equivalent in the classical sense although they came fairly close to meeting the standards.

The functional meaningfulness of the test forms as CR tests was studied by examining how they placed individuals with regard to their mastery of the tasks in the defined performance universe, and their use in evaluating the instructional techniques used to teach for the objectives. A Programmed Instruction (PI) text was developed for this purpose and was validated for the target population (i.e., the sample of the G Study) by following the steps of a systems model for development of instructional materials. The first formative evaluation consisted of one-on-one testing and an initial small group testing. The sample used for the second formative evaluation consisted of 23 persons drawn from the target population. The test forms developed earlier were administered to this sample as pretests and again as posttests in conjunction with the text. The resulting information was examined for agreement of the instruments regarding (1) the success of the individuals, and (2) the effectiveness of the text.

The instruments did not have high task-to-task agreement about the success of individuals. They agreed very well, however, about the overall performance of the sample with regard to the set of tasks as a whole in each test form. The relation to the agreement indices with the error Δ was investigated and the implications of the G theory are discussed in the paper. The agreement indices appeared, in the context of absolute decisions, to be analogous to correlation coefficients in the context of comparative decisions.

The gain indices for the PI text as yielded by the two test forms did not correspond task-to-task or person-to-person. They agreed very closely, however, with respect to the set of tasks over all persons in the sample.

The results of the study indicate that the G theory holds promise for CR testing. What size G coefficient is "acceptable" for CR tests, however, appears to be an issue that is to be resolved by future research findings.

Order No. 74-25,462, 214 pages.

COMPARATIVE PERFORMANCE OF BLACK AND WHITE TEENAGERS ON SELECTED TESTS OF AGGRESSIVE POTENTIAL

RUDOLPH, Pearlje Marietta Jessie, Ph.D.
The University of Michigan, 1974

Co-Chairmen: Dan H. Cooper, Alvin D. Loving

Program

The object of this study is to select and administer a set of measures regarding certain aspects of aggressive and hostile behavior and to analyze the results to determine possible patterns of differences according to race and sex.

Procedure

The subjects used in the study were drawn from two areas. The fifty blacks were students in a 10th grade physics class in Michigan. The fifty white teenagers were 10th grade students in a high school physics class in Illinois which has a similar socio-economic environment. The Guilford-Zimmerman Temperament Survey and the Rosenzweig Frustration Study tests were administered to both groups of teenagers. The tests were given on two different days in Illinois and the same procedures were followed at a high school in Michigan.

Among the one hundred subjects tested (14-16), the fifty whites were chosen to provide twenty-five white males and twenty-five white females, and the same procedure was followed with the black students in Michigan. All of the students were in a college preparation curriculum. The sample is representative of college bound students within two given areas, but seeking the same goals and curriculum.

Each subject was asked to read the instructions on the front of his booklet silently as the examiner read them aloud. The tests were scored by the examiner according to directions regarding aggression and reaction to aggression.

Major Findings

Investigation in the study was built around the following hypothesis: when taken from similar populations, black adolescents will show a higher degree of hostility than white adolescents.

In the testing, the white group responded with more hostility, as measured by the instruments designed to tap subconscious and conscious levels of consciousness in responses. Furthermore, results of this study indicate that sex differences were of major influence in the outcome. This was especially true with the black group. The black male was found least likely to express hostile impulses at a subjective level. Black boys expressed less hostility than white boys; and less than black girls. Conversely white boys expressed more hostility than either sex among blacks; and more hostility than white girls.

The findings may be related to patterns which are instilled in black boys, particularly, that develop the feelings and attitudes of being less wanted, less likely to be right, more dependent and less powerful, thus, he responds in a manner that is less likely to bring increased self-derogation and social reprisal; or it may be that black males of the type sampled are indeed more secure psychologically.

Conclusion

The findings could be due to the fact that the black male is turning his frustration inward; or he may be contented with his progress and himself and doesn't feel threatened. On the other hand, the white male might be just the opposite, either free to express his hostility, or subject to greater frustration, therefore he subjects others to hostile and aggressive behavior.

Order No. 74-25,314, 92 pages.

TEST ANXIETY REDUCTION—A COMPARISON OF THREE APPROACHES

SHOWALTER, John Michael, Ph.D.
The Ohio State University, 1974

Adviser: Professor Herman Peters

The purpose of this study was to compare three counseling methods of reducing test-taking anxiety among elementary and junior high school students. The three methods employed were Systematic Desensitization (S.D.), Rational-Emotive Therapy (RET) and Study-Skills Counseling (S.S.C.).

Considerable research exists to support the effectiveness of S.D. in reducing test-taking anxiety. Significant results have been obtained using such criterion measures as improvement in course grades, improvement in cumulative point-hour, change in achievement test scores, change on various self-report anxiety measures, and change in physiological conditions such as pulse, G.S.R., respiratory rate, etc. Fewer studies have been reported using RET to reduce test-taking anxiety, although those that are available support the effectiveness of RET as an appropriate treatment technique. Research using S.S.C. to reduce test-taking anxiety is also limited, but several studies do support its effectiveness compared to no-treatment control groups.

101 students were selected from a group of 161 elementary and junior high school students on the basis of scores on the Suinn Test Anxiety Behavior Scale (STABS). These students were assigned to one of three treatment conditions and a no-treatment control group. The first group received 6 treatment sessions of approximately one hour in length of systematic desensitization. Students were taught the principles of muscular relaxation and then constructed an anxiety hierarchy based upon familiar events prior to and during test-taking in their school. The second group received approximately 6-7 hours of group counseling, based on an RET framework. The third group received approximately 6-7 hours of S.S.C. where they participated in such activities as how-to-study lectures, practice in using the SQ3R method of study, practice in taking examinations, and discussions on types of examinations and handling problems that each present. The fourth group of students were assigned to a no-treatment control group who were tested prior to and following the administration of all treatment conditions. The treatments were administered by two experienced teachers who had some coursework in guidance and counseling.

It was hypothesized that each treatment condition would prove superior to the control group on the following three dependent criteria: (1) Scores on the reading part of the Stanford Achievement Test (S.A.T.); (2) Scores on the arithmetic part of the Stanford Achievement Test; and (3) Scores on the Modified Suinn Test Behavioral Anxiety Scale. These tests were administered pre and post treatment. In addition, all treatment groups were to be compared with each other on the same three dependent criteria to detect differences among treatment effectiveness.

Using an ANOV, significant differences were found on the Newman-Keuls post-hoc test between the following treatment and control groups, listed by dependent criteria: (S.A.T. Reading Test (RET exceeded the control group ($p < .01$); (S.A.T. Math Test (RET exceeded the control group ($p < .01$); S.D. exceeded the control group ($p < .01$), and S.D. exceeded S.S.C. ($p < .01$); (Suinn Test Anxiety Behavioral Scale) RET exceeded S.S.C. ($p < .01$); S.D. exceeded S.S.C. ($p < .05$); and the control group exceeded S.S.C. ($p < .01$).

The most impressive finding was on the arithmetic test, where both Systematic Desensitization and Rational-Emotive Therapy exceeded the control group at the .01 level. It was postulated that some of this difference might have been due to the narrow range of scores on this instrument; and, to the discouragement of the control group toward the end of the study. The unusual finding on the Suinn Test Anxiety Behavioral Scale (the control group mean score exceeded all treatment groups) may be explained in part by the variability of such an instrument and its questionable validity for measuring

test-taking anxiety. Recommendations include further research at the elementary grade levels, using more refined self-report measures and other criteria, such as physiological changes and behavioral observations by teachers and peers. Follow-up studies would help determine the permanency of any reduction in anxiety levels.

Order No. 74-17,807, 176 pages.

RISK-TAKING AND READING TEST PERFORMANCE. [To obtain a microfilm copy please order directly from the National Library of Canada at Ottawa]

CREALOCK, Carol Marie, Ph.D.
University of Toronto (Canada), 1973

A battery of tests measuring academic risk taking, reading achievement and attitudes towards school and risk was administered to 134 girls and 136 boys from seven, grade eight classes. The tests were given to groups of approximately thirty students in their own classrooms in the following order: (1) Attitude Scale, (2) Probability of Success Risk Taking, (3) Reading Comprehension Test (Power), (4) Word Knowledge Test (Guessing), (5) Word Information Risk Taking, (6) Picture Information Risk Taking, (7) Reading Comprehension Test (Speed), (8) Word Knowledge Test (No Guessing) from which both a reading score and the Multiple Choice Exam Risk Taking score were derived, and (9) Attitude toward the morning's tests. Total testing time was about three hours.

The major hypotheses stated that there would be a significant positive correlation among the four measures of academic risk taking, and that there would be a significant positive correlation between reading and risk taking.

The results suggest no support for a general academic risk taking tendency. The two measures of Information Risk Taking (Word and Pictures) did show a significant positive relationship ($r = .294$) but this correlation was very low given the similarity between the two measures. A low correlation was found between Word Information Risk Taking and Multiple Choice Exam Risk Taking ($r = .136$) but this correlation accounts for less than two percent of the total variance which does not suggest a stable risk taking trait.

Some reading measures correlated positively with Word Information Risk Taking and negatively with Probability of Success Risk Taking suggesting that guessing on a word with incomplete information and preferring a high degree of certainty of success on a choice dilemma problem are related to high scores on reading tests. In general however, reading test performance and risk taking did not yield consistently significant correlations and where a relationship was indicated, it accounted for less than five percent of the variance at best. Significant correlations between reading scores and risk taking more frequently occurred in females versus males and in poor readers versus good readers.

The attitude items did not correlate significantly with reading except for 'being compared to classmates.' This item was considered to suggest competition among fellow students. It was negatively rated by the total sample but the relationship to reading was strongest for poor readers. In this group, students who did not like competition scored higher on reading tests than did students who did not mind it.

The results do not suggest that academic risk taking is a consistent general trait within subjects that correlates positively with reading performance. However, before these hypotheses can be strongly rejected, further research into the area is needed.

CONCURRENT VALIDITY OF THE PEABODY INDIVIDUAL ACHIEVEMENT TEST WITH FOURTH-GRADE LEARNING DISABLED AND NON-LEARNING DISABLED CHILDREN USING THE METROPOLITAN ACHIEVEMENT TEST AS THE CRITERION MEASURE

BRAY, Nancy Ann Moreland, Ed.D.
East Texas State University, 1973

Adviser: Gerald W. Gattshall

Purpose of the Study: The purpose of the study was to investigate the concurrent validity of the Peabody Individual Achievement Test (PIAT) for learning disabled (LD) and non-learning disabled (Non-LD) fourth-grade children, using the Metropolitan Achievement Test (MAT) as the criterion measure. In addition, differences between correlations were explored for the variables of sex and ethnic identification.

Procedures: The population consisted of children enrolled in an experimental project designed to evaluate the effectiveness of three different teaching methods with learning disabled children in Dallas, Texas. There were 222 subjects: 107 LD and 115 Non-LD. The mean chronological age was 9 years, 8 months for the LD sample and 9 years, 7 months for the Non-LD sample. The mean Intelligence Quotient was 99 for the LD group and 106 for the Non-LD group, as measured by the California Short-Form Test of Mental Maturity.

Testing was counterbalanced so that approximately one-half of the LD and one-half of the Non-LD subjects were administered the PIAT followed by the MAT. The remaining subjects were administered the MAT followed by the PIAT.

Pearson product-moment coefficients of correlation were calculated between PIAT and MAT raw scores. In addition, Fisher's z was computed between comparable subtests to determine the significance of difference between correlations for the entire LD and Non-LD samples, subdivided by sex and ethnic identification.

Findings: Pearson r s ranged from .45 to .86 for comparable PIAT and MAT subtests for the entire LD sample. For the entire Non-LD sample, r s ranged from .56 to .77. No significant differences were found between correlations of comparable subtest scores between the entire LD and Non-LD samples, between the LD males and LD females, between the Non-LD males and Non-LD females, or among the Non-LD Caucasian, Negro, or Mexican-American subjects. For the LD group, however, when subdivided by ethnic identification, significant differences were found between the correlations of the PIAT Reading Recognition and MAT Total Reading scores for the Caucasian-Negro and Negro-Mexican-American comparisons.

Conclusions: It was concluded that substantial concurrent validity was demonstrated between the two instruments for both the LD and Non-LD samples, and when the samples were subdivided by sex and ethnic identification with one exception, for the PIAT subtests of Mathematics, Reading Recognition, Reading Comprehension, and Spelling. General Information and Total Test scores on the PIAT were not evaluated, since there were no comparable MAT subtests. Order No. 74-11,889, 158 pages.

AN INVESTIGATION OF A SCORING PROCEDURE DESIGNED TO ELIMINATE SCORE VARIANCE DUE TO GUESSING IN MULTIPLE-CHOICE TESTS

CROSS, Lawrence Hartwell, Ph.D.
University of Pennsylvania, 1973

Supervisor: Dr. James J. Diamond

The present study investigated a novel scoring procedure designed to eliminate or reduce score variance due to guessing on multiple-choice tests. A series of three teacher-written algebra tests was administered to 12 sections of eleventh-grade students attending a suburban Philadelphia high school. The six participating teachers agreed to use the scores from these tests for grading purposes, and the students were so informed. The examinees were directed to respond to each of the tests in two distinct ways: the Coombs response mode, which was used with appropriate directions during the first part of the testing period; and the conventional response mode, which was used with directions that encouraged guessing during the second part of the test

period. Two initial tests were designed to acquaint the students with the novel response mode and to provide feedback on their performance. Only the data from a third test were used for the experimental analysis. The use of these two response modes permitted an estimate of the number of items that an examinee answered correctly under the following conditions: without guessing; and with guessing limited to either two, three, or more choices. This provided a guessing-free score and a series of partially guessing-free score sets, which served as criteria representing different guessing-score components. The conventional-mode responses were scored, using number-right scoring, formula scoring, and the novel scoring procedure proposed in this study. The last procedure involved computing scores for each examinee on several variables: the proportion of correct choices marked, the mean difficulty and discrimination indices of all choices marked (both right and wrong), and the variability of difficulty of correctly marked choices. The square of these variables and the cross-product between each of these variables were then computed which yielded 27 predictor variables. These variables were then used in a forward-selection regression to predict the guessing-free scores, partially-guessing-free score sets, and two external validity measures. The results were double cross validated, and the correlation of the cross validated novel scores with each criterion was compared with the correlation of number-right scores and formula scores with the same criteria. In every case, the cross validated novel scores predicted the criterion less well than did the number-right or formula scores. An attempt was also made to predict several score sets based on only those items answered correctly through guessing. The results of these analyses provided no additional insight as to how the scoring variables related to guessing. A possible explanation for the relative inability of these variables to predict the criteria was offered. Several noteworthy score characteristics were also observed for various score sets.

Order No. 74-14,041, 115 pages.

RELATIONSHIPS OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN AND THE ILLINOIS TEST OF PSYCHOLINGUISTIC ABILITIES WHEN ADMINISTERED TO LEARNING DISABLED CHILDREN

Nell Willis PAGE, Ph.D.
Texas A&M University, 1973

Chairman: Dr. Donald G. Barker

The WISC and ITPA subtest scores of 150 learning disabled children were submitted to canonical analysis. Five canonical roots emerged. When the data were further analyzed on the basis of age classification, two roots were obtained for children under the age of eight years and four roots were obtained for children above the age of eight years. Differentiation of ability was evident with increase in age. This study also examined intercorrelations between subtest scores of the two instruments and between subtest scores within each instrument. The results indicated significant correlations, especially between three ITPA subtests and WISC verbal subtests. A multiple regression analysis was conducted to determine: (1) the predictive validity of ITPA subtests for the following criterion variables: Verbal Scale Score, Performance Scale Score, and Full Scale Score and (2) the predictive validity of WISC subtests for the ITPA Psycholinguistic Age Scores. All of the obtained validities were significant at the 0.01 level of confidence with the best predictive validity being ITPA subtest scores in predicting WISC Full Scale Scores. Implications of the findings of this investigation for the assessment of learning disabilities in children are given.

Order No. 74-1040, 106 pages.

THE USE OF THE MODIFIED HAND TEST AND PICTORIAL STUDY OF VALUES TO DIFFERENTIATE BETWEEN SUCCESSFUL AND UNSUCCESSFUL EDUCABLE MENTALLY RETARDED WORK-STUDY STUDENTS

Russell Allen HARDESTY, Ph.D.
The University of Oklahoma, 1973

Major Professor: Dr. P. T. Teska

This study was conducted to determine if the Modified Hand Test and Pictorial Study of Values could differentiate between successful and unsuccessful educable mentally retarded Work-Study students. A total of 50 subjects were individually administered the Modified Hand Test. The PSV was also administered to each school group. The subjects were enrolled in EMR Work-Study programs in eight high schools of the greater Oklahoma City, Oklahoma area. Twenty-five subjects were in each group. Statistical analysis was applied to each of the scoring categories of the Modified Hand Test and scale of value of the PSV to determine if significant differences existed between the groups. Median and quartile ranges were calculated for each scoring category of the Modified Hand Test and values of the PSV for both groups. The Median Test, with Yates correction for continuity, was used to test for significant differences between the successful and unsuccessful workers. The findings from the evaluations were: 1) Significant differences occurred between the successful and unsuccessful groups in the ACT ($p < .05$); MAL ($p < .05$); and $(1 - \frac{INT}{ENV})^2$ ($p < .05$). 2) Successful workers gave significantly more ACT responses on the female than on the male stimulus cards of the Modified Hand Test. 3) A significant difference between groups occurred for the Theoretical Scale of the PSV, with the successful workers scoring higher ($p < .05$).

The ACT, PATH and $(1 - \frac{INT}{ENV})^2$ of the Modified Hand Test and the six values and Genlike score of the PSV were correlated using Kendall's rank correlation, tau. A significant correlation was found between $(1 - \frac{INT}{ENV})^2$ and the Theoretical value ($p = .0047$).

Conclusions

Only three aspects of the Modified Hand Test: ACT, MAL, $(1 - \frac{INT}{ENV})^2$ successfully differentiated between the successful and unsuccessful workers.

Successful workers have more psychological energy available to achieve material goals as indicated by higher ACT scores. Early childhood experiences in which the successful worker was accepted probably accounts for the higher number of ACT responses on the female stimulus cards than on male and child cards.

The successful worker had fewer MAL responses which indicates less psychological stress and anxiety. The unsuccessful workers probably experience more doubts about their ability to succeed than do the successful workers. The successful workers show a more equal awareness of interpersonal relationship and environmental needs than do the unsuccessful workers as evidenced by the lower $(1 - \frac{INT}{ENV})^2$ of the successful worker. This awareness allows the successful worker to depend upon others for guidance on their job.

The awareness of the successful workers dependency on others is supported by their higher Theoretical scores on the PSV. The significant $(1 - \frac{INT}{ENV})^2$ - Theoretical correlation further supported this need.

The Modified Hand Test and PSV cannot be used exclusively to differentiate or predict successful and unsuccessful workers, but could be used in conjunction with experienced personnel to predict and recommend vocational training and habilitation

plans to best insure a successful training program and best provide the EMR Work-Study student job placement best suited to his needs.

Recommendations for Further Study

Little research has been done with the Modified Hand Test. The aspect of thirty stimulus cards and sex divisions of the cards provide numerous aspects for future studies.

The scoring of the Hand Test is somewhat too general and is difficult to score some items as a definite category. A research project in which the scoring system would be refined and made less ambiguous would be beneficial.

The $(1 - \frac{INT}{ENV})^2$ score could be an additional measure of the Hand Test with which to evaluate personality. It has the potential of presenting a more complete measure of adjustment than do the single scoring categories and deserves additional research.

A more extensive study might be conducted with the PSV to establish norms for use with adolescent EMR students. Also, the PSV might be refined to be more specifically suited for use with the EMR adolescent. The instructions are somewhat difficult for the EMR to understand even when read to him. The pictures are also presented in a confusing manner. Research could also be conducted with different instructions and presentations of the pictures.

There was no control of teacher-coordinator bias in this study. A study might be conducted in which the interpersonal relationship of the teacher-coordinator and Work-Study students affected the vocational success rating. A comparison of the ratings of teacher-coordinators, vocational counselors and employers of the Work-Study students could also be included in the study.

Order No. 73-31,474, 74 pages.

A COMPARISON OF A GROUP BENDER VISUAL MOTOR GESTALT TEST WITH THE INDIVIDUAL BENDER VISUAL MOTOR GESTALT TEST WITH KINDERGARTEN CHILDREN AND THE RELATIONSHIP OF THE TWO BENDER TESTS WITH IQ

William Edward CASKEY, JR., Ph.D.
Kent State University, 1973

Director: Louis A. Fliegler

The purpose of this investigation was to study the relationship between kindergarten pupils' performance on a group Bender Visual Motor Gestalt Test and on an individual Bender test. The study also sought to assess the relationship between Bender test results, when scored by the Koppitz system, and IQ. Focus was placed on the practice effect, which is defined as participation in one form of the Bender test administration (group or individual) and in the same or alternate form of the Bender test administered within a ten-day or two-week period. The effect of sex on Bender test scores was evaluated.

The instruments used were (a) the Otis-Lennon Mental Ability Test, Primary I level, Form J, 1967 revision; (b) the standardized individual Bender test; and (c) a group Bender test made of transparencies of the Bender cards.

The sample consisted of 193 kindergarten age children. Using a stratified random sampling method based upon Otis-Lennon IQ scores, Ss were assigned to one of the four test groups. Bender testing periods were from ten days to two weeks apart.

Following determination of inter-scorer reliability using the Koppitz scoring system ($r = .99$), all Bender protocols were scored by the investigator. Pearson product moment correla-

tions were then obtained to assess (a) the coefficient of equivalency between the group Bender test and the individual Bender test; and (b) the coefficient of stability for both the group and the individual Bender test. In addition, Pearson product moment correlations were run between group and individual Bender test scores from the first testing period and IQ and for the combination of all first testing period Bender scores and IQ. Where applicable, Fisher's r to Z transformations were run to determine the relationship between the obtained correlations. Analysis of variance was computed to determine if there were any significant differences between any of the groups with respect to (a) IQ, (b) Bender test scores for the first and second testing periods, and (c) Bender scores between testing periods. Expectancy tables were prepared to demonstrate the relationship between Bender test scores and IQ.

Tests of significance (t tests) were used to determine if there were any significant differences between the performances of boys and girls with respect to scores on the Bender tests or IQ scores.

The conclusions drawn from this investigation were:

- 1) Kindergarten children's performance on both forms of the Bender test is stable over a ten-day to two-week period.
- 2) In terms of the findings there is clear evidence of the equivalency of the two forms of administration of the Bender test when results are scored by the Koppitz system ($r = .87$).
- 3) The data from this study give clear evidence that the group and individual Bender tests correlate significantly with IQ (group $r = -.66$; individual $r = -.51$; combined group and individual $r = -.59$).
- 4) The data from this study give clear evidence that the Bender test, both group and individual, correlates significantly with IQ ($r = -.59$).
- 5) The use of the Bender test to differentiate between above average Otis-Lennon IQ scores and below average Otis-Lennon IQ scores was not supported.
- 6) While kindergarten girls perform significantly better on the Bender test than do kindergarten boys, the difference cannot, on the basis of the data from this study, be attributed to differences in IQ.

Order No. 73-32,337, 126 pages.

AN EVALUATION OF THE TORRANCE TESTS OF CREATIVE THINKING

Lilia da Rocha BASTOS, Ph.D.
University of Southern California, 1973

Chairman: Professor Michael

Purpose: The purpose of this study was to provide a critical analysis and evaluation of the Torrance Tests of Creative Thinking (TTCT) from evidence reported in its technical manual and in selected empirical studies against the criteria, or standards, set forth in Standards for Educational and Psychological Tests and Manuals, published in 1966 by the American Psychological Association. Twenty-six major ESSENTIAL criteria selected from the Standards constituted the key elements of a congruence model against which the TTCT was evaluated. The evaluative process which was guided by answering six questions constituting the research problem could be integrated within one main question: To what extent does the TTCT meet the ESSENTIAL criteria for (a) Dissemination of Information (two criteria—A1 and A2); (b) Interpretation (four criteria—B1, B2, B3, and B4); (c) Validity (seven criteria—C1, C2, C3, C4, C5, C6, and C7); (d) Reliability (six criteria—D1, D2, D3, D4, D5, and D6); (e) Administration and Scoring (two criteria—E1 and E2); and (f) Scales and Norms (five criteria—F1, F2, F4, F5, and F6)?

Methodology: A model of congruence evaluation constituted the methodology employed in this investigation. Evidence gathered from the TTCT technical manual and from empirical studies pertaining to the test was rated or judged as "satisfactory," "questionable," "unsatisfactory," or "not applicable," according to its compliance to the requirements stated in each of the 26 major ESSENTIAL criteria selected as the principal parameters or model for the evaluation of the test.

Findings: Results from the evaluation of the TTCT against each of the 26 major ESSENTIAL criteria in the Standards indicated the following judgments regarding the adequacy of the TTCT in meeting each criterion within each of the six major groupings of requirements: (a) Dissemination of Information—1 "satisfactory" and 1 "questionable"; (b) Interpretation—2 "satisfactory" and 2 "questionable"; (c) Validity—3 "satisfactory," 3 "questionable," and 1 "unsatisfactory"; (d) Reliability—4 "satisfactory," 1 "questionable," and 1 "not applicable"; (e) Administration and Scoring—2 "satisfactory"; and (f) Scales and Norms—3 "satisfactory," 1 "questionable," and 1 "unsatisfactory."

Conclusions: The following conclusions seemed to be warranted: (1) Since the TTCT has satisfactorily met only 15 (about 58 percent) of the ESSENTIAL criteria set forth in the Standards and has received eight "questionable" and two "unsatisfactory" ratings, the test cannot be considered as an entirely adequate instrument for generalized use. (2) Since it was in the division of Administration and Scoring in the Standards that the TTCT has satisfactorily met all the ESSENTIAL criteria against which it was evaluated, the major strength of the test would appear to reside in that category. (3) The major weaknesses of the TTCT would seem to be related to the category of Validity in which the test was assigned three "questionable" and one "unsatisfactory" ratings. (4) When the somewhat tentative character of measures purported to tap creativity is considered, the inadequacies observed in the TTCT may be hypothesized to reflect the general state of the art in the area of creativity.

Recommendations: The findings and conclusions evaluated in this study led to the following recommendations: (1) The deficiencies observed in the TTCT, represented by ratings of "questionable" and "unsatisfactory," should be noted and remedied to whatever extent possible. (2) Only when considerable caution is exercised should the TTCT be used for purposes other than exploratory research. (3) Other tests purporting to measure creative thinking, divergent production, or creativity also should be subjected to evaluations against the criteria set forth in the Standards so that prospective users could have a broader basis upon which to make their test selection.

Order No. 73-31,629, 101 pages.

THE STRUCTURE OF INTELLECT MODEL AS A BASIS FOR CROSS CULTURAL ANALYSIS OF TESTS

Herbert Leon WHITE, Ph.D.
New York University, 1973

Chairman: Professor Philip R. Merrifield

In this study two groups, one Black, one White, were compared on the manner in which a series of relations emerged among 29 variables applying to each group. Twenty-one of the variables were Structure-or-Intellect tests, seven were Civil Service sub-tests, and one was a job performance criterion. The relationships determined were: the factor structure of the 21 Structure-of-Intellect tests in each group; the relations of the Civil Service sub-tests to the factors found; and the relations of Structure-of-Intellect tests, Civil Service tests, and combinations of both, to criterion.

Most research studies in the United States which compare races and groups in level of performance on psychological tests do not deal adequately with the necessity to establish comparative validity, in the sense advocated by Irvine (1969), of the test used to measure possible differences. This current study examines the concepts of factor validity and comparative validity and their bearing on the problem of test results in different ethnic groups. The study is, in addition, directed toward meeting the needs for: devising job selection tests that will not discriminate unfairly against minority groups; gaining greater understanding of the growth and development of intellectual abilities; investigating the possible role of behavioral and "social intelligence" tests as job selection techniques; and de-

veloping a rationale for development of valid tests in Civil Service and other settings.

It was hypothesized that, for the two groups, the relations with the criterion would be different, the addition of information from the SI model would improve prediction, the factor structures would be different, some CS tests would relate differently to factors in each group, and some CS tests would correlate positively with SI factors.

The question as to whether the same factors previously established on an American White population would emerge in the Summer Supervisor population could not be answered definitely. Instead of the traditional SI factors in pure form, there appeared four factors which seemed to be composites of the qualities attributed to the hypothesized SI factors. The configuration of the SI marker tests loading on these factors suggests that further investigation of a torus SI model previously proposed by Varela might be profitably investigated as an alternative to the cubical model now hypothesized by Guilford.

While the four factors emerged in each of the groups analyzed, the comparative validity of SI tests varied between the two groups.

The hypothesis that prediction of criteria from the Civil Service tests would be different in the Urban Corps population than the Summer Supervisor population was supported by findings based on multiple regression techniques.

In both groups, SI tests tended to correlate with the criterion more than did the Civil Service tests.

The addition of information from the SI tests appeared to improve the prediction of criteria appropriate to the Summer Supervisor population, and similarly for the Urban Corps population.

While the factor components of these Civil Service Tests were partially revealed by the techniques used in this study, the Civil Service and the SI tests did not demonstrate comparative validity across the two groups studied. It is suggested that no tests be used to compare different cultural groups unless the tests have in addition to meeting other criteria, equivalent factor structures.

Further study is suggested of the SI model or an adaptation of such a model with respect to cross cultural studies, and better construction of selection tests.

Order No. 73-30,144, 166 pages.

test was used as a test of significance. Results revealed that not all teachers in the sample purported to use the same information in arriving at grades.

The large variations in predictive validities of standardized tests which occur when the course grade is used as the sole criterion leads to a second question: Does teacher consistency/inconsistency in assigning grades influence the predictive validity of standardized tests?

Grades awarded by the forty-two teachers and Language and Reading subtest scores of the Comprehensive Tests of Basic Skills earned by their students were gathered. After the grades and scores of the students had been grouped according to teacher, a Pearson product-moment coefficient of correlation between the final English grades and Comprehensive Tests of Basic Skills--Reading subtest raw scores was computed. The same procedure was used between the final English grades and the Comprehensive Tests of Basic Skills--Language subtest scores. Correlations between English grades and Comprehensive Tests of Basic Skills--Reading subtest scores were grouped according to the factor group to which the teacher had been assigned during the factor analysis process. Each group of correlations was averaged. Resulting correlations were tested for homogeneity and found significantly different at the .01 level. The same procedure was followed with correlations between the final English grade and the Comprehensive Tests of Basic Skills--Language and they, too, were found to be significantly different. It was concluded that there is a difference in information purportedly used by teachers in the determination of marks, and that there is a difference in correlations between standardized test scores and students' final marks when students' scores and grades have been grouped according to the type of information purportedly used by their teachers in assigning marks.

Order No. 74-1745, 98 pages.

A STUDY OF THE RELATIONSHIP BETWEEN SELECTED GRADING CRITERIA AND THE USE OF STANDARDIZED TEST SCORES AS PREDICTORS OF GRADES

Hale Gene DEXTER, Ph.D.
University of Missouri - Kansas City, 1973

This study investigated some of the factors which contribute to variability in grading and related these factors to the effectiveness of standardized tests in forecasting academic success in tenth grade English.

Standardized academic achievement and aptitude tests have shown substantially varied correlations when used to predict academic performance even when tests have demonstrated high coefficients of reliability. One reason for these variations might be the uncontrolled variation in grades themselves. Research has paid scant attention to this possibility.

Teachers commonly suggest that marks should be based on academic competence; however, subjective factors may be applied. This leads to a question: Do teachers use the same information in arriving at grades?

The Marking-Factor Inventory was administered to fifty-six teachers of tenth grade English in a large urban school district to determine what variables influence them in determining marks. Forty-two teachers responded.

A Spearman rank-order correlation between teachers was computed resulting in a 42 X 42 correlational matrix, which was factor analyzed. The Kolmogorov-Smirnov two-sample